



Comparison of the Performance of Machine Learning Algorithms Based on Artificial Intelligence with the State Space Model in Forecasting Tax Revenues of West Azerbaijan Province

Rahim Dabbagh* 

Associate Professor, Department of Industrial Engineering, Faculty of Mineral Industries and Technologies, Urmia University of Technology, Urmia, Iran.

Ahmad Ezzati Shourgoli 

Assistant Professor, Afagh Higher Education Institute, Urmia, Iran.

Raheleh Shabanpoor 

M.Sc. Student in Financial Engineering and Risk Management, Afagh Higher Education Institute, Urmia, Iran.

Abstract

Tax revenues are among the most significant sources of government income, comprising a substantial portion of the government budget. These revenues are essential for funding public services, implementing infrastructure projects, and covering administrative and operational expenses. Accurate tax revenue forecasting is a critical tool for economic planning and decision-making. Reliable forecasts enable governments to plan budgets effectively, formulate optimal tax policies, and enhance economic management by reducing uncertainties. This study uses monthly tax revenue data from West Azerbaijan Province spanning April 2005 to June 2024 to forecast tax revenues using machine learning algorithms, including Random Forest, Support Vector Regression, Lasso Regression, Deep Learning, and the State Space Model. The findings indicate that among the evaluated algorithms—Random Forest, Support Vector Regression, Lasso Regression, and Deep Learning—the Deep Learning model achieves the highest accuracy in both in-sample and out-of-sample forecasts based on RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) metrics. However, the State Space Model emerges as a robust tool for modeling and forecasting financial and economic variables, particularly those exhibiting seasonal,

* Corresponding Author: R.Dabbagh@uut.ac.ir

How to Cite: Dabbagh, R., Ezzati, A. & Shabanpoor, R. (2026). Comparison of the Performance of Machine Learning Algorithms Based on Artificial Intelligence with the State Space Model in Forecasting Tax Revenues of West Azerbaijan Province, *Journal of Tax Research*, 33(68), 125-169.

monthly, or daily patterns. Its ability to capture unobservable components, irregularities, and dynamic trends provides a notable advantage in analyzing complex economic data. A comparative analysis of the State Space Model as a robust econometric approach and Deep Learning as a cutting-edge machine learning method demonstrates that the State Space Model outperforms Deep Learning in forecasting tax revenues for West Azerbaijan Province.

Introduction

Tax revenues play a pivotal role in the fiscal management of governments, providing essential funding for public services, infrastructure development, and economic stabilization. Accurate forecasting of tax revenues is crucial for ensuring effective budgeting and minimizing the risks associated with fiscal deficits or inefficient resource allocation. However, given the inherently dynamic and non-linear nature of economic systems—especially at the regional level—traditional econometric models often fall short in capturing complex patterns of revenue fluctuation.

In recent years, the rapid advancement of artificial intelligence and machine learning techniques has opened new frontiers in economic forecasting. Particularly, deep learning and hybrid machine learning approaches have shown promising potential in improving predictive accuracy by modeling non-linear, non-stationary, and irregular data behaviors. While deep learning algorithms such as neural networks and fuzzy logic systems can effectively capture hidden relationships in financial datasets, structural time-series models like state-space frameworks remain powerful in accounting for unobserved components such as trends, cycles, and seasonal variations.

This study investigates the comparative performance of several artificial intelligence-based machine learning algorithms—including Random Forest, Support Vector Regression, LASSO regression, and fuzzy-based deep learning—alongside the traditional state-space model in forecasting monthly tax revenues for West Azerbaijan Province from 2005 to 2024. The objective is to assess not only the in-sample forecasting accuracy but also the robustness of each model in out-of-sample prediction scenarios, particularly in the context of economic policy shifts and irregular revenue shocks. By doing so, the research aims to offer data-driven insights into the applicability of modern AI techniques in regional fiscal planning and contribute to the optimization of public finance strategies in developing economies.

Methods and Material

This study applies a comparative modeling approach to forecast tax revenues in West Azerbaijan Province using a monthly dataset spanning from April 2005 to June 2024. The methodology integrates traditional econometric models with

multiple machine learning algorithms to assess their effectiveness in both in-sample and out-of-sample forecasting. West Azerbaijan Province using a monthly dataset spanning from April 2005 to June 2024. The methodology integrates traditional econometric models with multiple machine learning algorithms to assess their effectiveness in both in-sample and out-of-sample forecasting.

Data Collection and Preprocessing

The dataset comprises monthly tax revenue data collected from official provincial records, encompassing 231 observations. It includes diverse economic phases, policy shifts, and structural changes in tax administration. To ensure data quality, missing observations were addressed, and normality of distribution was tested using the Jarque-Bera statistic, which indicated non-normality. Given this, machine learning models that do not require distributional assumptions were favored. The data was also normalized prior to model training to enhance convergence and mitigate scale-related distortions.

Model Selection and Development

Five predictive models were implemented for performance comparison:

1. **Random Forest (RF):** A robust ensemble learning method that constructs multiple decision trees and aggregates their predictions to improve accuracy and reduce overfitting.
2. **Support Vector Regression (SVR):** A supervised learning model that utilizes kernel functions to handle non-linear relationships in time-series data.
3. **Standard Machine Learning with Lasso Regression:** A linear regression model enhanced with L1 regularization for feature selection and simplification.
4. **Deep Learning Model (DLM):** A multi-layer neural network incorporating fuzzy logic, designed for capturing non-linear, non-stationary patterns in tax revenue trends.
5. **State-Space Model:** A structural time-series model capable of capturing unobservable components such as trend, seasonality, and stochastic volatility using Kalman filtering.

Each model was trained using in-sample data from April 2005 to September 2023. Forecasts for the period October 2023 to June 2024 were evaluated as out-of-sample predictions.

Performance Evaluation Metrics

To evaluate the forecasting performance, two key error metrics were used:

- **Root Mean Square Error (RMSE):** Quantifies the standard deviation of

prediction errors, giving more weight to larger deviations.

- **Mean Absolute Percentage Error (MAPE):** Measures the average absolute percent deviation of forecasts from actual values, allowing comparison across models and timeframes.

Computational Environment

All models were developed using R (for Random Forest, SVR, and Lasso) and MATLAB (for Deep Learning). The Kalman filter-based state-space model was implemented using specialized econometric packages (Ox-Metrics).

This methodology enables a robust evaluation of whether advanced AI-driven models—especially deep learning architectures—can outperform traditional methods in capturing complex dynamics in tax revenue data and respond effectively to economic shifts and irregularities.

The empirical analysis conducted in this study compares the performance of five forecasting models—Random Forest (RF), Support Vector Regression (SVR), Lasso Regression, Deep Learning (DL), and the State Space Model (SSM)—in predicting monthly tax revenues for West Azerbaijan Province. The results are evaluated based on both in-sample (April 2005 to September 2023) and out-of-sample (October 2023 to June 2024) forecasts using RMSE and MAPE as performance metrics.

Results and Discussion

In-Sample Forecasting Results

Among the machine learning models, the Deep Learning algorithm exhibited the best performance in in-sample forecasting with the lowest RMSE (150,245 million IRR) and MAPE (9.88%). This demonstrates the model’s capacity to capture intricate patterns and nonlinearities in the time series. The standard machine learning model using Lasso Regression closely followed, while Random Forest and SVR recorded relatively higher prediction errors, with SVR showing the weakest in-sample performance due to overfitting and sensitivity to kernel parameters.

Table 1. In-Sample Prediction Accuracy of Different Models for Four Machine Learning Models

Model	RMSE (Million IRR)	MAPE (%)
Random Forest	186,486	10.13%
Support Vector Regression	323,619	23.46%
Standard Machine Learning	171,245	9.90%
Deep Learning	150,245	8.88%

Source: Research findings.

Out-of-Sample Forecasting Results

In the out-of-sample evaluation, the Deep Learning model maintained its superiority among machine learning algorithms with an RMSE of 850,173 million IRR and a MAPE of 11.21%. This performance suggests that the model generalizes well to unseen data, likely due to its multi-layer structure and use of fuzzy inference to handle uncertainty. However, it was observed that certain months, particularly March and June 2024, exhibited significant discrepancies between predicted and actual values, reflecting the challenges posed by irregular fiscal events or policy changes.

Table 2. Comparison of Out-of-Sample Forecasting Accuracy for Four Machine Learning Models

Model	RMSE (Million IRR)	MAPE (%)
Random Forest	1,156,901	15.58%
Support Vector Regression	1,732,153	21.55%
Standard Machine Learning	1,197,739	15.64%
Deep Learning	850,173	11.21%

Source: Research findings.

Comparative Performance of State Space Model

Despite the promising performance of Deep Learning, the State Space Model ultimately delivered the most accurate forecasts overall. It achieved lower RMSE and MAPE values in out-of-sample predictions compared to all machine learning models. The model’s structure, which explicitly accounts for unobservable components such as trend, seasonality, and stochastic volatility, provided it with a decisive advantage in modeling the underlying dynamics of tax revenue data. Its adaptability through Kalman filtering allowed for real-time adjustments in the presence of unexpected fluctuations, a feature that traditional machine learning models often lack.

Discussion

These results underscore the importance of model selection based on the nature of the economic data being analyzed. While Deep Learning offers substantial flexibility and power in capturing non-linear relationships, its performance may deteriorate in cases of structural breaks or limited data. The State Space Model, by contrast, is particularly effective in financial time-series environments with strong seasonal components and policy-driven shocks. The comparative findings suggest that a hybrid approach—combining the

feature extraction capabilities of deep learning with the structural strength of econometric models—may yield optimal forecasting performance. This has practical implications for policymakers seeking reliable tools for fiscal planning and revenue management in volatile economic contexts.

Conclusion

Accurate forecasting of tax revenues is vital for effective fiscal policy implementation and long-term economic planning. This study evaluated and compared the performance of various machine learning algorithms—including Random Forest, Support Vector Regression, Standard Machine Learning (Lasso), Deep Learning, and the State Space Model—in predicting monthly tax revenues in West Azerbaijan Province from April 2005 to June 2024.

The results indicate that among the machine learning models, the Deep Learning algorithm achieved the highest accuracy in both in-sample and out-of-sample forecasting, reflecting its ability to model complex, non-linear patterns in financial time series. Nevertheless, the State Space Model ultimately outperformed even the Deep Learning model in terms of predictive accuracy during out-of-sample evaluation, particularly in handling irregular trends, seasonal dynamics, and unobservable components such as structural breaks.

These findings suggest that while AI-driven methods offer significant improvements in predictive performance, traditional econometric models like the State Space framework remain powerful tools in economic forecasting—especially when dealing with non-stationary or seasonally affected data. A hybrid strategy that leverages the strengths of both approaches could offer a more comprehensive and adaptable forecasting system.

For future research, integrating additional variables such as macroeconomic indicators, policy reforms, and real-time transaction data may further enhance model robustness. Moreover, the incorporation of advanced techniques like ensemble deep learning, attention-based models, or reinforcement learning could provide deeper insights and stronger predictive capabilities in complex fiscal environments.

In conclusion, this study highlights the practical benefits of combining machine learning and econometric modeling in tax revenue forecasting, providing valuable guidance for policymakers and fiscal planners aiming to improve the accuracy and reliability of revenue projections.

Keywords: West Azerbaijan Province, Artificial Intelligence (AI), Deep Learning, Tax Revenue Forecasting.

JEL Classification: B23, K34, H2.



سازمان امور مالیاتی کشور

-- مجله علمی، پژوهشنامه مالیات --

شماره ۶۸، دوره ۳۳، زمستان ۱۴۰۴، ۱۶۹-۱۲۵

taxjournal.ir

DOI: 10.66224/taxjournal.34.68.125

مقایسه عملکرد الگوریتم‌های یادگیری ماشین مبتنی بر هوش مصنوعی با الگوی فضا حالت در پیش‌بینی درآمدهای مالیاتی استان آذربایجان غربی

دانشیار گروه مهندسی صنایع، دانشکده صنایع و فناوری‌های معدنی، دانشگاه صنعتی ارومیه، ارومیه، ایران.

رحیم دباغ* 

استادیار موسسه آموزش عالی آفاق، ارومیه، ایران.

احمد عزتی شورگلی 

کارشناسی ارشد رشته مهندسی مالی و مدیریت ریسک، موسسه آموزش عالی آفاق، ارومیه، ایران.

راحله شعبانپور 

چکیده

پیش‌بینی روند متغیرهای مالی و اقتصادی از جمله درآمدهای مالیاتی ابزار مهمی برای تصمیم‌گیری و برنامه‌ریزی اقتصادی است. این پیش‌بینی‌ها به دولت‌ها کمک می‌کنند تا بودجه خود را به درستی برنامه‌ریزی و سیاست‌های مالیاتی بهینه را تعیین کنند. همچنین می‌تواند به بهبود مدیریت اقتصادی و کاهش عدم قطعیت‌ها منتج شود و دولت را در برنامه‌ریزی مالی بلندمدت یاری کند. بر این اساس در مقاله حاضر با استفاده از داده‌های ماهانه درآمدهای مالیاتی استان آذربایجان غربی طی دوره زمانی فروردین ۱۳۸۴ الی خرداد سال ۱۴۰۳ و با استفاده از الگوریتم‌های یادگیری ماشین مبتنی بر جنگل تصادفی، رگرسیون بردار پشتیبان، یادگیری ماشین مبتنی بر رگرسیون لاسو، یادگیری عمیق ماشین و الگوی فضا حالت به پیش‌بینی درآمدهای مالیاتی استان آذربایجان غربی پرداخته شده است. نتایج تحقیق نشان داد که در بین الگوریتم‌های یادگیری ماشین، الگوی یادگیری عمیق ماشین بر اساس معیارهای میانگین مجذور خطا و میانگین قدر مطلق درصد خطا بهترین عملکرد را در پیش‌بینی درون نمونه‌ای و برون نمونه‌ای از درآمدهای مالیاتی استان آذربایجان غربی دارد. با این حال الگوی فضا حالت نیز به عنوان یک ابزار قدرتمند در مدل‌سازی و پیش‌بینی متغیرهای مالی و اقتصادی است، به خصوص زمانی که روندهای فصلی، ماهانه و یا روزانه در آن متغیر وجود داشته باشد. همچنین وجود اجزاء غیر قابل مشاهده، اجزاء نامنظم و وجود شیب تصادفی در مدل‌سازی الگوهای مالی و اقتصادی از ویژگی‌های بارز این الگو است. با توجه به حجم داده‌های مورد استفاده در تحقیق حاضر و برای دوره زمانی ذکر شده، مقایسه عملکرد الگوی فضا حالت به عنوان یک ابزار قدرتمند در الگوهای اقتصادسنجی و الگوی یادگیری عمیق ماشین به عنوان یکی از ابزارهای قوی و مهم در یادگیری ماشین نشان داد که الگوی فضا حالت عملکرد بهتری حتی نسبت به یادگیری عمیق ماشین در پیش‌بینی درآمدهای مالیاتی استان آذربایجان غربی دارد.

کلیدواژه‌ها: استان آذربایجان غربی، هوش مصنوعی، یادگیری عمیق ماشین، پیش‌بینی درآمدهای مالیاتی.

طبقه‌بندی JEL: B23، K34، H2.

* نویسنده مسئول: R.Dabbagh@uut.ac.ir

مقدمه

اهمیت مالیات در تأمین بودجه و نقش ابزاری آن جهت دستیابی به اهداف کلان اقتصادی موجب می‌شود تا با انجام پیش‌بینی‌های دقیق‌تر بتوان دولت را در تدوین برنامه‌ها و رسیدن به اهداف خود یاری رساند. با توجه به نظام مالیاتی موجود و رویه‌های اجرایی، میزان تحقق درآمدهای مالیاتی مصوب، می‌تواند نقش بسیار مهمی در برنامه‌ریزی‌های اقتصادی دولت داشته باشد (Fikriya & Hikmawati, 2020). مدیریت چیزی جز عمل تصمیم‌گیری نیست و مهمترین فاکتور برای تصمیم‌گیری، پیش‌بینی وضعیت آینده در تمام حوزه‌های مرتبط با برنامه‌ریزی بوده است. منظور از پیش‌بینی مالیاتی، میزان مالیات وصول شده و مقدار وجوه یک دولت است که در یک دوره زمانی مشخص دریافت می‌شود. تغییرات و پیچیدگی‌های زیادی که در دهه اخیر در بسیاری از سیستم‌های اقتصادی و اجتماعی ایجاد شده است، منجر به ایجاد پیش‌بینی با دقت و اطمینان کم شده است (Molapo et al., 2019). بنابراین شناخت، پیش‌بینی و کنترل رفتار این پدیده‌ها، درک رفتار گذشته آنها و چگونگی تغییرات این رفتارها در دوره‌های مختلف ضروری است. به طور کلی می‌توان گفت که پیش‌بینی عبارت است از برآورد پیشامدهای آینده و هدف از آن، کاهش ریسک در تصمیم‌گیری است. از آنجایی که پیش‌بینی همیشه با مقداری خطا همراه است، بنابراین جهت کاهش آن با داشتن اطلاعات بیشتر در مورد سیستم و استفاده از بازه پیش‌بینی این مشکل را تا حدی می‌توان حل کرد (Abdi et al., 2011). پس از ارائه اولین مدل‌های کلان‌سنجی و گسترش این نوع مدل‌ها، مطالعات اقتصادسنجی فراوانی توسط دولت‌ها، بانک‌های مرکزی و سایر ارگان‌های تصمیم‌گیر در کشورهای مختلف به منظور پیش‌بینی و ارزیابی سیاست‌ها مورد استفاده قرار گرفت (Febriminanto & Wasesa, 2022). از سوی دیگر، با توسعه روش‌های مدل‌سازی، الگوهای دیگری مانند مدل‌های موسوم به تعادل عمومی قابل محاسبه، مدل‌های سری زمانی، الگوهای یادگیری ماشین و هوش مصنوعی به عنوان ابزارهای پیش‌بینی مطرح شد؛ اما نزدیک به دو دهه است که رویکردهای ترکیبی در پیش‌بینی مطرح شده است. در این رویکردها ترکیبی از چند روش به منظور کاهش خطای پیش‌بینی استفاده می‌شود (Maulia & Sofyan, 2018).

مطالعات متعددی در حوزه پیش‌بینی با روش‌های مختلف آماری، اقتصادسنجی و روش‌های یادگیری ماشین انجام گرفته است، اما مطالعات بسیار کمی در ایران در حوزه

پیش‌بینی درآمدهای دولت بالاخص درآمدهای مالیاتی انجام شده است، لذا با توجه به اهمیت بسیار زیاد روند درآمدهای مالیاتی جهت برنامه‌ریزی و پیش‌بینی روند کلی درآمدهای دولت، ضرورت دارد که پیش‌بینی قابل اعتمادی از روند این منبع درآمدی دولت انجام شود، لذا بر این اساس و با به کارگیری الگوریتم‌های یادگیری ماشین مبتنی بر هوش مصنوعی، در مقاله حاضر درآمدهای مالیاتی استان آذربایجان غربی پیش‌بینی شده است.

در ادامه این مقاله و در بخش دوم، مروری بر ادبیات تحقیق، در بخش سوم، روش انجام تحقیق و معرفی داده‌ها، در بخش چهارم، نتایج تجربی تحقیق و در بخش پنجم، نتیجه‌گیری و پیشنهادها ارائه شده است.

مبانی نظری و پیشینه پژوهش

عوامل مؤثر در تعیین مقدار درآمدهای مالیاتی

در یک تقسیم‌بندی کلی، مقدار وصولی درآمدهای مالیاتی را می‌توان از دیدگاه کلان اقتصادی و یا از دیدگاه مودیان مورد بررسی قرار داد که در ادامه، عوامل کلان اقتصادی و عواملی که بر رفتار مودیان در پرداخت یا عدم پرداخت مالیات اثرگذار هستند، تشریح می‌شود.

دیدگاه کلان اقتصادی

افزایش و کاهش درآمدهای مالیاتی دولت تحت تأثیر عوامل متعددی قرار دارد که می‌توان آن‌ها را از دیدگاه اقتصادی تحلیل کرد. این عوامل به طور کلی به دو دسته عوامل داخلی (درون‌زا) و عوامل خارجی (برون‌زا) تقسیم می‌شوند. در ادامه، به بررسی این عوامل و تفسیر آنها پرداخته می‌شود:

عوامل مذکور به طور مستقیم و غیر مستقیم بر درآمدهای مالیاتی دولت تأثیر می‌گذارند. سیاست‌های مالیاتی مؤثر و اجرای قوی قوانین مالیاتی می‌تواند به افزایش درآمدهای مالیاتی کمک کند. در عین حال، رشد اقتصادی، سطح اشتغال و سرمایه‌گذاری نیز عوامل کلیدی در تأمین منابع مالی دولت از طریق مالیات‌ها هستند. ساختار اقتصادی و تنوع آن می‌تواند به پایداری درآمدهای مالیاتی کمک کند، در حالی که وابستگی به چند بخش محدود می‌تواند این درآمدها را ناپایدار کند.

شرایط اقتصادی جهانی و روابط بین‌المللی نیز نقش مهمی در درآمدهای مالیاتی دارند. تغییرات در قیمت‌های جهانی کالاها، رونق یا رکود اقتصادی جهانی و تحولات سیاسی می‌توانند تأثیرات چشمگیری بر درآمدهای مالیاتی داشته باشند. در نتیجه، تحلیل دقیق این عوامل و تدوین سیاست‌های مناسب می‌تواند به دولت‌ها در بهبود و افزایش درآمدهای مالیاتی کمک کند (Kalaš et al., 2020).

عوامل داخلی (درون‌زا)

سیاست‌های مالیاتی

ساختار نرخ‌های مالیاتی: نرخ‌های مالیاتی بالا می‌توانند موجب افزایش درآمدهای مالیاتی شوند، اما در عین حال ممکن است انگیزه‌های اقتصادی را کاهش دهند و منجر به فرار مالیاتی شوند. نرخ‌های پایین‌تر می‌توانند انگیزه‌های اقتصادی را تقویت کنند، اما ممکن است به کاهش درآمدهای مالیاتی منجر شوند.

گستره پایه مالیاتی: افزایش تعداد افراد و شرکت‌های مشمول مالیات و کاهش معافیت‌ها می‌تواند به افزایش درآمدهای مالیاتی منجر شود.

اجرای قوانین مالیاتی: تقویت سیستم‌های جمع‌آوری و اجرای مالیات و کاهش فرار مالیاتی می‌تواند به افزایش درآمدهای مالیاتی کمک کند (Milenković et al., 2020).

فعالیت‌های اقتصادی

رشد اقتصادی: افزایش تولید ناخالص داخلی (GDP) به افزایش درآمدها و سود شرکت‌ها منجر می‌شود که به تبع آن درآمدهای مالیاتی نیز افزایش می‌یابد.

سطح اشتغال: افزایش سطح اشتغال و کاهش بیکاری منجر به افزایش درآمدهای حاصل از مالیات بر درآمد افراد می‌شود.

سرمایه‌گذاری: افزایش سرمایه‌گذاری در بخش‌های مختلف اقتصادی می‌تواند به افزایش تولید و درآمدها و در نتیجه افزایش درآمدهای مالیاتی منجر شود (Pradhan et al., 2024).

ساختار اقتصادی

تنوع اقتصادی: اقتصادی که به چندین بخش مختلف وابسته است، درآمدهای مالیاتی پایدارتری دارد. وابستگی به یک یا دو صنعت می‌تواند درآمدهای مالیاتی را ناپایدار کند.

بخش غیر رسمی اقتصاد: کاهش حجم بخش غیر رسمی اقتصاد می تواند درآمدهای مالیاتی را افزایش دهد زیرا فعالیت های اقتصادی بیشتری تحت پوشش نظام مالیاتی قرار می گیرند (Andrejovska & Pulikova, 2018).

عوامل خارجی (برونزا)

شرایط اقتصادی جهانی

رکود و رونق جهانی: شرایط اقتصادی جهانی بر تقاضای کالاها و خدمات صادراتی کشور و درآمدهای حاصل از مالیات های تجاری تأثیر می گذارد. نوسانات قیمت کالاها: تغییرات قیمت کالاهای اساسی مانند نفت، گاز و مواد معدنی می تواند درآمدهای مالیاتی کشورهای که به صادرات این کالاها وابسته هستند را تحت تأثیر قرار دهد (Neog & Gaur, 2020).

روابط بین الملل

تجارت بین المللی: افزایش تجارت بین المللی و کاهش موانع تجاری می تواند به افزایش درآمدهای مالیاتی از طریق افزایش حجم تجارت و واردات منجر شود. تغییرات در قوانین بین المللی مالیاتی: تغییرات در توافق نامه های مالیاتی بین المللی و همکاری های بین المللی در مبارزه با فرار مالیاتی می تواند بر درآمدهای مالیاتی تأثیر بگذارد (Karpowicz & Majewska, 2018).

تحولات سیاسی و اجتماعی

ثبات سیاسی: ثبات سیاسی و اجتماعی می تواند به ایجاد محیطی مناسب برای رشد اقتصادی و افزایش درآمدهای مالیاتی منجر شود. تحولات جمعیتی: تغییرات در ساختار جمعیتی، مانند افزایش تعداد نیروی کار یا مهاجرت، می تواند بر درآمدهای مالیاتی تأثیر بگذارد (Castro & Camarillo, 2014).

روش‌های پیش‌بینی

روش‌های کلان اقتصادی

مدل‌های اقتصاد کلان به تحلیل تأثیرات کلی سیاست‌های مالی و اقتصادی بر درآمدهای مالیاتی می‌پردازند. استفاده از مدل‌های اقتصادی، مانند مدل‌های تعادل عمومی پویای تصادفی^۱ و مدل‌های نهادی و ساختاری، به همراه تحلیل هزینه-فایده، می‌تواند به دولت‌ها در پیش‌بینی دقیق‌تر درآمدهای مالیاتی و اتخاذ تصمیمات مؤثرتر در زمینه سیاست‌های مالیاتی کمک کند. این روش‌ها به دولت‌ها امکان می‌دهند تا تأثیرات کوتاه‌مدت و بلندمدت تغییرات سیاست‌های مالیاتی را بر اقتصاد و درآمدهای مالیاتی تحلیل و ارزیابی کنند، و به این ترتیب به مدیریت بهینه منابع مالی و برنامه‌ریزی بودجه کمک کنند (Arfaee et al., 2022).

مدل‌های اقتصادسنجی

الگوهای رگرسیونی

این مدل به بررسی رابطه خطی بین یک متغیر مستقل مثلاً تولید ناخالص داخلی و یک متغیر وابسته (درآمدهای مالیاتی) می‌پردازد. فرمول ساده این مدل به صورت $Y = \alpha + \beta X + \epsilon$ است، که در آن Y درآمد مالیاتی، متغیر مستقل X تولید ناخالص داخلی، α عرض از مبدا، β ضریب شیب و ϵ خطای تصادفی است. با استفاده از داده‌های تاریخی تولید ناخالص داخلی و درآمدهای مالیاتی، می‌توان ضریب شیب و عرض از مبدا را تخمین زد و سپس درآمدهای مالیاتی آینده را پیش‌بینی کرد. این مدل به بررسی رابطه بین چندین متغیر مستقل مانند تولید، نرخ بیکاری، تورم و یک متغیر وابسته (درآمدهای مالیاتی) می‌پردازد.

فرمول این مدل به صورت $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$ است، که در آن Y درآمد مالیاتی و X_i متغیرهای مستقل مختلف هستند. این مدل به تحلیل تأثیر همزمان چندین عامل اقتصادی بر درآمدهای مالیاتی کمک می‌کند و پیش‌بینی دقیق‌تری ارائه می‌دهد (Ma et al., 2019).

الگوهای خودرگرسیون میانگین متحرک

این مدل‌ها به تحلیل داده‌های سری زمانی و پیش‌بینی مقادیر آینده بر اساس روندهای گذشته

1. DSGE

می‌پردازند. مدل خودرگرسیون میانگین متحرک شامل سه بخش اصلی است: اتورگرسیون، درجه انباشتگی و میانگین متحرک. برای پیش‌بینی درآمدهای مالیاتی، می‌توان داده‌های تاریخی را به مدل خودرگرسیون میانگین متحرک اعمال کرد و الگوهای گذشته را برای پیش‌بینی مقادیر آینده استفاده کرد.

توسعه و گسترش الگوهای خودرگرسیون میانگین متحرک

مدل‌های خودرگرسیون میانگین متحرک به دلیل انعطاف‌پذیری و قابلیت تطبیق با داده‌های سری زمانی مختلف، به طور گسترده‌ای در تحلیل و پیش‌بینی متغیرهای اقتصادی و مالی استفاده می‌شوند. هر یک از مدل‌های پیشرفته خودرگرسیون میانگین متحرک قابلیت‌های خاص خود را دارند و برای کاربردهای مختلف مناسب هستند. انتخاب مدل مناسب بستگی به ماهیت داده‌ها و هدف پیش‌بینی دارد. مدل خودرگرسیون میانگین متحرک چرخشی مارکوف برای داده‌های با تغییرات رژیمی، مدل خودرگرسیون میانگین متحرک واریانس شرطی برای داده‌های با نوسانات متغیر، مدل خودرگرسیون میانگین متحرک فضا حالت برای سیستم‌های دینامیک پیچیده و مدل خودرگرسیون میانگین متحرک غیر خطی برای داده‌های با رفتارهای پیچیده و غیر خطی مناسب هستند. استفاده از این مدل‌ها می‌تواند به دقت بیشتر پیش‌بینی‌ها و تحلیل‌های سری زمانی کمک کند (Holikova et al., 2023). از جمله این الگوها می‌توان به الگوهای خودرگرسیون میانگین متحرک غیر خطی، خودرگرسیون میانگین متحرک چرخشی مارکوف، خودرگرسیون میانگین متحرک فضا حالت، خودرگرسیون میانگین متحرک با لحاظ ناهمسان واریانس شرطی و ... اشاره کرد.

الگوی داده‌های تابلویی

این مدل‌ها از داده‌های ترکیبی زمانی-مقطعی استفاده می‌کنند که شامل داده‌های چندین واحد (مانند کشورهای مختلف) در طول زمان است.

روش‌های ریاضی

روش‌های ریاضی در پیش‌بینی درآمدهای مالیاتی و تعیین سیاست‌های بهینه از اهمیت بالایی برخوردارند. این روش‌ها شامل مدل‌های بهینه‌سازی و مدل‌های دینامیکی هستند که به تجزیه

و تحلیل و مدل‌سازی پیچیدگی‌های سیستم‌های اقتصادی می‌پردازند. روش‌های ریاضی، بهینه‌سازی و مدل‌های دینامیکی ابزارهای قدرتمندی برای تحلیل و پیش‌بینی درآمدهای مالیاتی و تعیین سیاست‌های مالیاتی بهینه فراهم می‌کنند. این روش‌ها به دولت‌ها کمک می‌کنند تا تصمیم‌گیری‌های مبتنی بر داده‌ها و تحلیل‌های دقیق داشته باشند و بهبود مدیریت اقتصادی و مالیاتی را فراهم کنند. استفاده از این روش‌ها می‌تواند به دقت بیشتر در پیش‌بینی‌ها و کارایی بالاتر در سیاست‌گذاری‌های مالیاتی منجر شود (Altaf et al., 2021).

روش‌های یادگیری ماشین و یادگیری ترکیبی در پیش‌بینی درآمدهای مالیاتی استفاده از روش‌های ترکیبی و تکنیک‌های یادگیری ماشین به‌طور چشمگیری دقت و کارایی پیش‌بینی درآمدهای مالیاتی را افزایش داده است. این روش‌ها توانایی تحلیل حجم زیادی از داده‌ها و شناسایی الگوهای پیچیده را دارند که روش‌های سنتی قادر به انجام آن‌ها نیستند. روش‌های ترکیبی و یادگیری ماشین ابزارهای قدرتمندی برای پیش‌بینی درآمدهای مالیاتی فراهم می‌کنند. این روش‌ها می‌توانند به دولت‌ها و سازمان‌های مالیاتی کمک کنند تا با دقت بیشتری برنامه‌ریزی کنند، سیاست‌های مالیاتی بهینه‌تری اتخاذ کنند و به بهبود مدیریت اقتصادی و مالیاتی دست یابند. استفاده از این تکنیک‌ها به معنای بهره‌گیری از توانمندی‌های فناوری‌های جدید و روش‌های پیشرفته تحلیل داده است که می‌تواند به نتایج بهتر و کارایی بالاتر در پیش‌بینی درآمدهای مالیاتی منجر شود (Hajej et al., 2017). از جمله الگوریتم‌های مهم یادگیری ماشین در رگرسیون می‌توان به رگرسیون بردار پشتیبان، رگرسیون لجستیک، رگرسیون مبتنی بر جنگل تصادفی، رگرسیون مبتنی بر درخت‌های تصمیم‌گیری، شبکه‌های عصبی مصنوعی و ... اشاره کرد.

پیشینه پژوهش

عباسیان و همکاران (۱۳۹۰) با استفاده از داده‌های مالیاتی ایران طی سال‌های ۱۳۵۰ الی ۱۳۸۵ با معادله تفکیک اجزاء خطا به منظور تحلیل منشأ خطای پیش‌بینی درآمدهای مالیاتی ایران پرداخته‌اند. محققین در مطالعه خود به این نتیجه دست یافتند که به‌طور متوسط در طول دوره مورد بررسی برآورد مالیات بر درآمد، مالیات بر ثروت، مالیات بر واردات و مالیات بر مصرف و فروش بالاتر از حد واقعی (خوشبینانه) و برآورد مالیات بر اشخاص حقوقی پایین‌تر از حد واقعی

(محافظه کارانه) بوده است. همچنین خطای پیش‌بینی مالیات بر اشخاص حقوقی، مالیات بر درآمد، مالیات بر واردات و مالیات بر مصرف و فروش به‌طور عمده از نوع غیر سیستماتیک و خطای پیش‌بینی مالیات بر ثروت بیشتر از نوع سیستماتیک بوده است.

نصراللهی و شاکر اردکانی (۱۳۹۵) با استفاده از داده‌های ایران طی سال‌های ۱۳۵۲ الی ۱۳۹۰ و با به کارگیری معادلات رگرسیون‌های به ظاهر نامرتبط به پیش‌بینی درآمدهای مالیاتی ایران پرداخته‌اند. محققین در مطالعه خود به این نتیجه دست یافتند که تولید ناخالص داخلی بدون نفت و نرخ ارز بر روی پیش‌بینی همه درآمدهای دولت تأثیرگذار بوده، ولی متغیرهای نرخ تورم و نرخ بیکاری به ترتیب، تنها در پیش‌بینی طراحان بودجه از درآمد مالیاتی و درآمد حاصل از فروش کالاها و خدمات مؤثر بوده است.

روحی و همکاران (۱۳۹۶) با استفاده از داده‌های ۳۰۰۰ شرکت بزرگ در ایران طی سال‌های ۱۳۸۸ الی ۱۳۹۴ و با به کارگیری زنجیره مارکوف به پیش‌بینی وصولی مالیات بر درآمد شرکت‌ها پرداخته‌اند. محققین در مطالعه خود به این نتیجه دست یافتند که دقت پیش‌بینی مدل زنجیره مارکوف با اختلاف کمی بهتر از تحلیل طیف منفرد بوده است. همچنین از مالیات قطعی شده هر سال، ۳۰ درصد در سال پس از آن و مابقی در دو سال و بیشتر بعد از آن وصول می‌شود.

زارعی و همکاران (۱۳۹۸) در مطالعه خود با استفاده از الگوریتم ژنتیک به بررسی پیش‌بینی اثر مالیات سبز بر مصرف و شدت انرژی پرداخته‌اند. محققان نشان دادند که بر اساس نتایج شبیه‌سازی توسط الگوریتم ژنتیک، چنان‌چه مالیات سبز افزایش یابد کاهش مصرف انرژی و در مقابل کاهش تولید را در پی خواهد داشت. همچنین اگر شدت انرژی به صورت نسبت مصرف انرژی به تولید در نظر گرفته شود، این افزایش مالیات سبز کاهش شدت انرژی را به همراه دارد. بنابراین نتایج نشان می‌دهد که مالیات سبز منجر به کاهش مصرف و شدت انرژی در ایران می‌شود.

دباغ و همکاران (۱۳۹۸) در مطالعه مقایسه عملکرد مالی بانک‌های اسلامی و بانک‌های متعارف در کشورهای منتخب نشان داده‌اند که بانک‌های اسلامی در برخی زمینه‌ها مانند مقاومت در برابر بحران‌های مالی عملکرد بهتری دارند، در حالی که بانک‌های متعارف در بهره‌وری عملیاتی و سودآوری کوتاه‌مدت قوی‌تر ظاهر می‌شوند. این نتیجه می‌تواند در تحلیل سیاست‌های مالیاتی و تصمیم‌گیری‌های نظارتی به ویژه در طراحی چارچوب‌های مالیاتی منصفانه و متناسب با ماهیت هر نوع بانکداری مؤثر باشد.

سعادت مهر (۱۴۰۰) با استفاده از داده‌های ایران طی سال‌های ۱۳۵۷ الی ۱۳۹۶ و با به کارگیری الگوی خودرگرسیون با وقفه‌های توزیعی به بررسی تأثیر متغیرهای کلان اقتصاد بر درآمدهای مالیاتی دولت در ایران پرداخته است. محقق در مطالعه خود به این نتیجه دست یافت که متغیرهای تولید سرانه بدون نفت، اشتغال و درآمدهای نفتی تأثیر مثبت و نرخ تورم تأثیر منفی بر درآمدهای مالیاتی دولت دارند. تولید سرانه بدون نفت بیشترین و نرخ تورم کمترین تأثیر را بر درآمدهای مالیاتی دارند.

عزتی شورگلی و همکاران (۱۴۰۲) به پیش‌بینی اثر وصل شدن تراکنش دستگاه‌های کارتخوان به سیستم امور مالیاتی در میزان افزایش وصولی سازمان امور مالیاتی استان آذربایجان غربی پرداخته‌اند. محققین با استفاده از داده‌های ماهانه درآمدهای مالیاتی استان طی سال‌های ۱۳۸۴ الی ۱۴۰۱ و با به کارگیری سه الگوی فضا حالت، چرخشی مارکوف و آریما نشان داده‌اند که مالیات ستانی مبتنی بر عملکرد تراکنش دستگاه‌های کارتخوان، درصد تحقق کل درآمدهای مالیاتی استان آذربایجان غربی در سال ۱۴۰۱ را تقریباً ۳۲ درصد افزایش داده است. دباغ و همکاران (۱۴۰۴) در پژوهشی به شناسایی چالش‌های اصلی پیاده‌سازی فناوری‌های صنعت ۴.۰ در صنایع کوچک و متوسط پرداخته‌اند. نتایج نشان داد که کمبود زیرساخت‌های فناوری اطلاعات، نبود نیروی متخصص و عدم آگاهی از سیاست‌های مرتبط از جمله موانع اصلی در بهره‌گیری از فناوری‌های نوین این صنایع هستند. از منظر مالیاتی، این یافته‌ها اهمیت حمایت‌های مالیاتی هدفمند از این صنایع را روشن می‌سازد؛ به گونه‌ای که سیاست‌های معافیت یا تخفیف مالیاتی می‌توانند انگیزه لازم برای سرمایه‌گذاری در زیرساخت‌ها، آموزش منابع انسانی و توسعه فناوری‌های دیجیتال را فراهم کرده و مسیر حرکت به سوی صنعتی‌سازی پایدار را هموارتر کنند.

بوتر و کادر^۱ (۲۰۱۰) در مطالعه خود بر اساس داده‌های کشورهای OECD به پیش‌بینی درآمدهای دولت شامل درآمدهای مالیاتی در کشورهای مذکور با استفاده از روش پیش‌بینی پسماندها پرداخته‌اند. محققین در مطالعه خود نشان دادند که بر اساس مقایسه روش‌های پیش‌بینی، تفاوت در مقدار پیش‌بینی می‌تواند تا حد زیادی به تفاوت در زمان پیش‌بینی‌ها و نوع مالیات بستگی داشته باشد. همچنین روش‌ها و عوامل مورد استفاده در پیش‌بینی، نتیجه را تا حد زیادی تحت تأثیر قرار می‌دهد.

کایرالا^۱ (۲۰۱۲) در پژوهشی با عنوان پیش‌بینی درآمد دولت در نپال با ارائه مدل‌های مختلف و مقایسه نتایج حاصل از پیش‌بینی آنها با استفاده از شاخص میانگین درصد خطا و میانگین قدر مطلق درصد خطا به این نتیجه رسیده که مدل‌های ARIMA درآمدهای دولت در نپال را با دقت بیشتری پیش‌بینی نموده است.

سویم و همکاران^۲ (۲۰۱۴) جهت پیش‌بینی متغیرهای اقتصادی ترکیه مبتنی بر شبکه‌های عصبی مصنوعی (ANN)، درخت تصمیم‌گیری و رگرسیون لاجیت را توسعه داده‌اند. نتایج مدل‌سازی و پیش‌بینی از روش‌های مبتنی بر یادگیری ماشین و هوش مصنوعی نشان داد که شبکه‌های عصبی مصنوعی بهترین عملکرد را در تجزیه و تحلیل خود داشتند.

چاتزیز و همکاران^۳ (۲۰۱۸) به بررسی و پیش‌بینی بحران مالی در بازار سهام با استفاده از رویکردهای مختلف یادگیری ماشین و همچنین یادگیری عمیق پرداختند. نتایج مطالعه مذکور نشان داد که رویکرد شبکه‌های عصبی مصنوعی عمیق به نسبت سایر روش‌ها در وضعیت بهتری در پیش‌بینی قرار دارد.

مالیکوف و همکاران^۴ (۲۰۲۱) در مطالعه خود بر اساس داده‌های فصلی کشور ازبکستان به پیش‌بینی درآمدهای مالیاتی با استفاده از مدل‌های سری زمانی پرداخته‌اند. محققین نشان داده‌اند که همگامی پیش‌بینی‌های بلندمدت از درآمدهای مالیاتی قابل اعتماد است، اگر سری‌های زمانی کافی و طولانی از آن داده‌ها وجود داشته باشد. افق پیش‌بینی باید با دوره‌ای مطابقت داشته باشد که طی آن روندها شناسایی شده و پارامترهای دینامیک برآورد می‌شوند. در این راستا، استفاده از روش‌های تحلیل اقتصادسنجی سنتی توسعه‌یافته برای برنامه‌ریزی مالیاتی بلندمدت غیر ممکن است. برای پیش‌بینی‌های کوتاه‌مدت، اطلاعات لازم در دسترس است و می‌تواند در برنامه‌ریزی‌های دولتی سودمند باشد.

لاهی‌ری و یانگ^۵ (۲۰۲۲) در مطالعه خود بر اساس داده‌های فصلی ایالت نیویورک در کشور آمریکا به پیش‌بینی درآمدهای مالیاتی با استفاده از تکنیک رگرسیون داده‌های با تواتر متفاوت پرداخته‌اند. محققین در مطالعه خود نشان دادند که درآمدهای مالیاتی ایالت نیویورک با یک مدل فرکانس ترکیبی با استفاده از تکنیک یادگیری ماشین قابل پیش‌بینی است.

1. Koirala (2012)

2. Sevim et al. (2014)

3. Chatzis et al. (2018)

4. Malikov et al. (2021)

5. Lahiri and Yang (2022)

طیب و همکاران^۱ (۲۰۲۳) در مطالعه خود بر اساس داده‌های مالیاتی کشور هند به پیش‌بینی درآمدهای مالیاتی با استفاده از تکنیک مدل‌های سری زمانی و شبکه عصبی مصنوعی پرداخته‌اند. محققین در مطالعه خود از الگوهای خودرگرسیون میانگین متحرک، روش اجزاء روند فصلی، هموارسازی ملایم و شبکه عصبی مصنوعی بهره جسته‌اند و نشان دادند که روش‌های غیر خطی مبتنی بر الگوهای خودرگرسیون میانگین متحرک در پیش‌بینی درآمدهای مالیاتی کشور هند نسبت به روش‌های خطی عملکرد بهتری دارد.

باتیستون و همکاران^۲ (۲۰۲۴) روشی را برای پیاده‌سازی بهینه سیاست‌های عمومی که شامل پیش‌بینی رفتار یا ویژگی‌های فردی است، ارائه داده‌اند. آنها با مرتبط ساختن خطاهای پیش‌بینی هر مدل، معیاری ساده برای رتبه‌بندی و انتخاب مدل بهینه فراهم کردند. این معیار به عنوان تفاوت بین رفاه اجتماعی ناشی از یک سیاست معین و رفاه اجتماعی حاصل از سیاستی بدون خطا تعریف شد، روش مذکور بر پایه پیش‌بینی اظهارنامه‌های نادرست مالیاتی که توسط افراد خوداشتغال و شرکت‌های انحصاری در ایتالیا صادر شده‌اند، به کار گرفته شد. نتایج نشان داد که این رویکرد می‌تواند منجر به افزایش قابل توجهی در درآمدها شود و مدل‌های جنگل تصادفی، علاوه بر ارائه پیش‌بینی‌های نسبتاً دقیق، دیدگاه‌های مهمی را نیز فراهم می‌کنند.

روش انجام تحقیق و معرفی داده‌ها

روش انجام تحقیق

جهت پیش‌بینی درآمدهای مالیاتی استان آذربایجان غربی از ۴ الگوریتم یادگیری ماشین شامل، جنگل تصادفی، رگرسیون بردار پشتیبان، یادگیری استاندارد ماشین مبتنی بر رگرسیون لاسو و یادگیری عمیق ماشین و همچنین الگوی فضا-حالت استفاده شده است تا نتایج پیش‌بینی آنها با هم مقایسه شود. در ادامه هر کدام از روش‌های مذکور تشریح می‌شود.

الگوی ساختاری اجزاء غیر قابل مشاهده مبتنی بر فضا حالت

در این الگو متغیرهای توضیحی تابعی از زمان هستند و با پارامترهای (اجزای غیر قابل مشاهده) در طول زمان به‌طور تصادفی تغییر می‌کنند. به عبارت دیگر، اجزای متفاوت-روند، فصلی،

1. Tayyib et al. (2023)

2. Battiston et al. (2024)

سیکلی و بی‌قاعده- سیر تکاملی سری زمانی را که به‌طور سنتی به‌صورت معین نشان داده شده‌اند، توصیف می‌کنند. مدل سری زمانی ساختاری روشی انعطاف‌پذیر است که می‌تواند مؤلفه‌های غیر قابل مشاهده را در مدل‌های رگرسیونی پویا وارد کند (Salami & Mafi, 2018). این مدل یک ابزار مناسب برای لحاظ این مؤلفه‌ها و برآورد روند ضمنی متغیر اقتصادی از جمله درآمدهای مالیاتی است. فرم کلی این الگو برای درآمدهای مالیاتی به شکل زیر است.

$$TAX_t^{STATE-SPACE} = \mu_t + \gamma_t + \varphi_t + \sum_i^p a_i TAX_{t-i} + \varepsilon_t \quad (1)$$

در رابطه (۱) μ_t متغیر روند، γ_t جزء فصلی یا ماهانه، φ_t جزء چرخه، TAX_{t-i} وقفه‌های درآمدهای مالیاتی استان آذربایجان غربی و ε_t جزء اخلاص الگو است. البته متغیر روند به صورت تصادفی در رابطه (۲) تصریح می‌شود.

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad (2)$$

$$\beta_t = \beta_{t-1} + \vartheta_t \quad (3)$$

معادله‌های ۲ و ۳ به ترتیب سطح روند و شیب روند را نشان می‌دهند. این فرآیند این گونه تفسیر می‌شود که روند این دوره با روند دوره قبل به اضافه شوک غیر قابل پیش‌بینی برابر است. با توجه به اینکه مدل مورد استفاده در مطالعه دارای جزء غیر قابل مشاهده است نمی‌توان روش‌های حداقل مربعات را برای تخمین به کار برد. با این وجود، وقتی که معادله ۱ همراه با معادلات ۲ و ۳ با هم تشکیل یک سیستم پویا را می‌دهند، می‌توان این سیستم پویا را به شکل فضای حالت نمایش داد که نمایشی از حالت یک سیستم با اجزای غیر قابل مشاهده است که می‌تواند به کمک فرآیند فیلتر شدن به‌روز شود. این سیستم در شکل فضای حالت به‌صورت دو معادله مجزا، یکی معادله انتقال و دیگری معادله اندازه‌گیری تنظیم می‌شود. ارائه یک سیستم پویا به شکل فضای حالت این امکان را فراهم می‌کند تا با کمک فیلتر کالمن، پارامترهای مجهول (ابریارامترها و سایر پارامترها) از طریق روش حداکثر راستمایی برآورد شوند. فیلتر کالمن یک دسته معادلات بازگشتی تولید می‌کند که با در نظر گرفتن یک حالت ابتدایی، به تخمین ضرایب مدل در زمان بعد می‌پردازد. این امر باعث می‌شود تغییرات درآمدهای مالیاتی که ناشی از تغییر سیاست‌های اقتصادی، پیشرفت تکنولوژی و فناوری در حوزه مالیات‌ستانی و یا ساختار اقتصادی است، در مراحل بعد تعدیل شود (Shiranifakhr et al., 2014).

الگوریتم‌های یادگیری ماشین

یادگیری ماشین مبتنی بر درخت‌های تصمیم‌گیری

جنگل تصادفی یک الگوریتم محبوب یادگیری ماشین است که به تکنیک یادگیری نظارت‌شده تعلق دارد. این یک روش یادگیری جمعی است که از کیسه‌گذاری (نمونه‌برداری بوت‌استرپ) استفاده می‌کند، به طوری که در زمان آموزش تعدادی درخت تصمیم‌گیری ایجاد کرده و خروجی آن کلاسی است که بیشترین تکرار را دارد (طبقه‌بندی) یا میانگین پیش‌بینی درخت‌های جداگانه (رگرسیون). این الگوریتم می‌تواند برای هر دو مسئله طبقه‌بندی و رگرسیون در یادگیری ماشین مورد استفاده قرار گیرد. با این حال، همچنین می‌تواند در تحلیل و پیش‌بینی سری‌های زمانی، چه داده‌های تک‌متغیره و چه چندمتغیره، با ایجاد متغیرهای تاخیری و متغیرهای فصلی و ماهانه استفاده شود.

یادگیری ماشین مبتنی بر جنگل تصادفی

جنگل تصادفی^۱ یک الگوریتم یادگیری ماشین است که برای طبقه‌بندی^۲ و رگرسیون^۳ استفاده می‌شود. این الگوریتم از روش ترکیب مدل‌ها^۴ بهره می‌برد. جنگل تصادفی از ترکیب تعدادی درخت تصمیم‌گیری^۵ به وجود می‌آید تا دقت پیش‌بینی‌ها را افزایش دهد و مشکلاتی مانند بیش‌برازش^۶ را کاهش دهد (Xu & Kong, 2024). الگوریتم جنگل تصادفی بر اساس دو اصل کلیدی عمل می‌کند: ۱- روش بوت‌استرپ^۷ و ۲- انتخاب تصادفی ویژگی‌ها^۸.

در روش بوت‌استرپ، الگوریتم به‌جای استفاده از کل مجموعه داده برای آموزش هر درخت تصمیم‌گیری، نمونه‌های بوت‌استرپ از داده‌ها می‌گیرد. نمونه‌های بوت‌استرپ شامل انتخاب تصادفی داده‌ها با جایگزینی است. به این ترتیب، هر درخت تصمیم‌گیری با مجموعه‌ای متفاوت از داده‌ها آموزش می‌بیند (Murorunkwere et al., 2023). در انتخاب تصادفی ویژگی‌ها، هر گره از درخت تصمیم‌گیری، به جای استفاده از تمامی ویژگی‌ها برای تقسیم داده‌ها، با یک زیرمجموعه تصادفی از ویژگی‌ها انتخاب می‌شود. این امر باعث می‌شود که درختان تصمیم‌گیری متنوع‌تر شوند و همبستگی میان آن‌ها کاهش یابد (Xu & Kong, 2024).

1. Random Forest

2. Classification

3. Regression

4. Ensemble Learning

5. Decision Trees

6. Overfitting

7. Bootstrap Sampling

8. Random Feature Selection

مراحل الگوریتم جنگل تصادفی

ایجاد نمونه‌های بوت‌استرپ: تعداد مشخصی از نمونه‌ها (معمولاً به اندازه مجموعه داده اصلی) با جایگزینی از مجموعه داده اصلی انتخاب می‌شوند.

ساخت درختان تصمیم‌گیری: برای هر نمونه بوت‌استرپ، یک درخت تصمیم‌گیری ساخته می‌شود. در هر گره از درخت، به جای استفاده از تمامی ویژگی‌ها، یک زیرمجموعه تصادفی از ویژگی‌ها انتخاب می‌شود. تقسیم داده‌ها در هر گره بر اساس بهترین ویژگی از میان این زیرمجموعه انجام می‌شود.

پیش‌بینی: برای طبقه‌بندی نتایج تمامی درختان تصمیم‌گیری جمع‌آوری شده و کلاس با بیشترین تعداد رأی به عنوان خروجی نهایی انتخاب می‌شود، برای رگرسیون میانگین پیش‌بینی‌های تمامی درختان تصمیم‌گیری به عنوان خروجی نهایی در نظر گرفته می‌شود (Xu & Kong, 2024).

یادگیری ماشین مبتنی بر رگرسیون بردار پشتیبان

رگرسیون بردار پشتیبان^۱ یا (SVR) یکی از تکنیک‌های قدرتمند یادگیری ماشین است که بر اساس اصول مشابه با ماشین‌های بردار پشتیبان برای طبقه‌بندی عمل می‌کند. این الگوریتم برای مسائل رگرسیون طراحی شده است و می‌تواند برای پیش‌بینی مقادیر پیوسته مانند درآمدهای مالیاتی استفاده شود. این الگوریتم به‌طور خاص در مواردی که داده‌ها دارای پیچیدگی بالا و الگوهای غیر خطی هستند، عملکرد خوبی دارد. الگوریتم رگرسیون بردار پشتیبان یکی از تکنیک‌های پیشرفته یادگیری ماشین است که به دلیل توانایی در مدل‌سازی روابط پیچیده و غیر خطی، برای پیش‌بینی درآمدهای مالیاتی بسیار مناسب است. با استفاده از این الگوریتم، می‌توان به پیش‌بینی‌های دقیقی از درآمدهای مالیاتی دست یافت که به تصمیم‌گیری‌های مالی و اقتصادی کمک شایانی می‌کند (Li-Xia et al., 2011).

اصول اولیه رگرسیون بردار پشتیبان

رگرسیون بردار پشتیبان هدفش یافتن یک تابع خطی در فضای ویژگی‌ها است که بتواند با حداقل خطا، داده‌های آموزش را پیش‌بینی کند. این الگوریتم از یک تابع خطی (یا غیر خطی) استفاده می‌کند و تلاش می‌کند تا مقدار پیش‌بینی شده (y) را تا حد امکان به مقادیر واقعی نزدیک

1. Support Vector Regression

کند، در حالی که خطای پیش‌بینی از یک مقدار مجاز (ϵ) بیشتر نشود (Zhang et al., 2023).

تابع خطی در رگرسیون بردار پشتیبان

فرض کنید داده‌ها به صورت (x_i, y_i) برای $i = 1, 2, \dots, n$ باشند، که بردار ویژگی‌ها y_i مقدار واقعی است. هدف SVR پیدا کردن تابع خطی رابطه (۴) است:

$$f(x) = w \cdot x + b \quad (۴)$$

که در آن w بردار وزن‌ها و b یک بایاس است (Zhang et al., 2023).

مفهوم حاشیه و مقدار مجاز خطا (ϵ)

در SVR، هدف این است که پیش‌بینی‌های $f(x_i)$ در فاصله‌ای به اندازه ϵ از مقادیر واقعی y_i قرار بگیرند. این مفهوم به شکل رابطه (۵) بیان می‌شود.

$$\epsilon \geq |y_i - f(x_i)| \quad (۵)$$

همچنین برای داده‌هایی که خارج از این حاشیه قرار می‌گیرند، جریمه‌هایی در نظر گرفته می‌شود (Li-Xia et al., 2011).

یادگیری عمیق ماشین مبتنی بر استنتاج فازی

یک شبکه عصبی مانند مغز انسان ساختار یافته است و از نورون‌های مصنوعی تشکیل شده است که با عنوان «گره» نیز شناخته می‌شود. این گره‌ها در سه لایه در کنار هم چیده شده‌اند: لایه ورودی، لایه (های) مخفی و لایه خروجی (Huang et al., 2007). داده‌ها، اطلاعاتی را در قالب ورودی به هر گره ارائه می‌دهند. گره، ورودی‌ها را در وزن‌های تصادفی ضرب می‌کند، آن‌ها را محاسبه کرده و یک بایاس به آن اضافه می‌کند. در نهایت، توابع غیر خطی، که به آن‌ها «توابع فعال سازی» گفته می‌شود، برای تعیین اینکه کدام نورون شلیک کند، اعمال می‌شوند. شبکه عصبی، تقریب محاسباتی مغز انسان است که از گراف بردارها و لبه‌ها برای نمایش عصب‌ها (نورون‌ها) و سیناپس‌ها استفاده می‌کند. شبکه با مدل‌سازی متغیرهای ورودی به عنوان لایه‌بردارها انجام می‌شود و سپس اوزان را به هر اتصال در گراف نسبت می‌دهد، در حالی که

سایر بردارها در سطوح متمایز گذاشته می‌شوند که حاکی از فاصله‌شان از گره‌های ورودی است (Kuan & White, 1994).

در حالی که الگوریتم‌های یادگیری عمیق دارای بازنمایی‌های خودآموزا هستند، این الگوریتم‌ها به شبکه‌های عصبی مصنوعی وابسته‌اند که نحوه محاسبه اطلاعات توسط مغز را منعکس می‌کنند. در طول فرایند آموزش، الگوریتم‌ها از عناصر ناشناخته در توزیع ورودی برای استخراج ویژگی‌ها، گروه‌بندی اشیاء و کشف الگوهای مفید داده‌ها استفاده می‌کنند. مشابه آموزش ماشین‌ها برای خودآموزی، این امر در سطوح مختلف، با استفاده از الگوریتم‌هایی برای ساخت مدل‌ها رخ می‌دهد. مدل‌های یادگیری عمیق از چندین الگوریتم استفاده می‌کنند (Nosratabadi et al., 2020).

الگوریتم‌های یادگیری عمیق و سیستم‌های فازی از دو حوزه مجزا در هوش مصنوعی و یادگیری ماشین هستند که هر یک مزایا و قابلیت‌های خاص خود را دارند. ترکیب این دو به صورت "یادگیری عمیق ماشین فازی" می‌تواند به پیش‌بینی دقیق‌تر و موثرتر درآمدهای مالیاتی کمک کند. در این سیستم‌ها، یادگیری عمیق توانایی مدل‌سازی روابط پیچیده و غیرخطی در داده‌ها را فراهم می‌آورد، در حالی که منطق فازی می‌تواند با عدم قطعیت‌ها و ابهامات در داده‌ها به خوبی کار کند (Lin et al., 2012).

منطق فازی به جای استفاده از مقادیر دقیق و قطعی، از مجموعه‌های فازی و مقادیر احتمالی برای مدل‌سازی داده‌ها و روابط بین آن‌ها استفاده می‌کند. در یک سیستم فازی، تابع عضویت برای هر مقدار، یک درجه عضویت نسبت داده می‌شود که نشان‌دهنده درجه عضویت آن مقدار در یک مجموعه فازی است. قوانین فازی، از قوانین "اگر-آنگاه" (IF-THEN) برای تعیین خروجی‌های سیستم استفاده می‌کند (Ran, 2021).

معیارهای ارزیابی پیش‌بینی الگو

معیارهای عمده‌ای برای ارزیابی پیش‌بینی الگوهای اقتصادسنجی و الگوریتم‌های یادگیری ماشین وجود دارد که این معیارها بر مبنای مقایسه مقادیر واقعی (Y_t) و مقادیر پیش‌بینی ($\hat{Y}_t^f$) قرار دارند. حال بر اساس الگوی فضا حالت و الگوریتم‌های یادگیری ماشین توضیح داده شده در قسمت قبلی، مقدار برازش شده و پیش‌بینی شده از این الگوها، تحت عنوان TAX_t^f است که بر اساس دو معیار زیر مورد ارزیابی قرار می‌گیرند.

میانگین ریشه مجذور خطا^۱ (MSE)

در رابطه (۶) طول دوره پیش‌بینی، TAX_t مقدار درآمدهای مالیاتی وصول شده در دوره مورد بررسی و TAX_t^f مقدار پیش‌بینی درآمدهای مالیاتی بر اساس یکی از الگوریتم‌های یادگیری ماشین و یا الگوی فضا حالت است.

$$RMSE = \sqrt{\frac{\sum_{t=T+1}^{T+M} (TAX_t^f - TAX_t)^2}{m}} \quad (۶)$$

میانگین قدر مطلق درصد خطا^۲ (MAPE)

معیار فوق متاثر از واحد اندازه‌گیری Y است، بدین معنی که بزرگ و کوچک بودن مقادیر Y موجب بزرگ و کوچک شدن معیار می‌شود. به این منظور معیار دیگری به نام میانگین قدر مطلق درصد خطا (MAPE) معرفی می‌شود که تحت تأثیر واحد اندازه‌گیری Y قرار ندارد (Gorgini et al., 2012).

$$MAPE = \frac{100 \sum_{t=T+1}^{T+M} \left| \frac{TAX_t^f - TAX_t}{TAX_t} \right|}{m} \quad (۷)$$

در معادلات (۶ و ۷)، TAX_t درآمدهای مالیاتی واقعی و وصول شده استان آذربایجان غربی طی دوره ۱۳۸۴ الی ۱۴۰۳ (به صورت ماهانه) است. TAX_t^f نیز مقادیر پیش‌بینی شده از الگوهای فضا حالت و یادگیری ماشین به دست آمده‌اند. لذا با جایگذاری هریک به جای TAX_t^f در معادلات بالا، خطای پیش‌بینی برای الگوهای برآوردی محاسبه خواهد شد. قابل ذکر است که هرچه خطای پیش‌بینی براساس معیارهای ارائه شده در رابطه (۷)، کوچک باشد، الگو قدرت پیش‌بینی بهتری نسبت به سایر الگوها دارد.

داده‌های تحقیق

داده‌های این مطالعه به صورت ماهانه و برای استان آذربایجان غربی و در بازه‌ی زمانی فروردین سال ۱۳۸۴ الی خرداد سال ۱۴۰۳ هستند. لازم به ذکر است که جهت پیش‌بینی درون‌نمونه‌ای از داده‌های دوره زمانی فروردین سال ۱۳۸۴ الی شهریور سال ۱۴۰۲ استفاده شده است و سپس با پیش‌بینی خارج از نمونه و برای ۹ ماه آتی شامل پیش‌بینی مهر سال ۱۴۰۲ الی خرداد سال ۱۴۰۳ به عنوان پیش‌بینی برون‌نمونه‌ای در نظر گرفته شده است.

1. Mean Squared Error

2. Mean Absolute Percentage Error

آمار توصیفی متغیر درآمدهای کل مالیاتی به صورت ماهانه بصورت خلاصه در جدول ۱ نشان داده شده است. با توجه به جدول مذکور، میانگین، میانه و انحراف معیار کل درآمدهای ماهیانه مالیاتی به ترتیب برابر با ۹۶۸،۸۰۸ میلیون ریال، ۳۹۷،۶۰۷ میلیون ریال و ۱،۴۸۸،۲۲۲ میلیون ریال طی فروردین سالهای ۱۳۸۴ الی خرداد ۱۴۰۳ است. مطابق با آماره آزمون جبارک برا و احتمال آزمون، توزیع متغیرها نرمال نیست.

داده‌هایی که توزیع نرمال ندارند، می‌توانند با استفاده از الگوریتم‌های یادگیری ماشین مدل‌سازی و پیش‌بینی شوند. بسیاری از الگوریتم‌های یادگیری ماشین نیازی به فرض توزیع خاص برای داده‌ها ندارند. برخی الگوریتم‌های یادگیری ماشین به توزیع داده حساس نیستند و می‌توانند مستقیماً بر روی داده‌های غیر نرمال استفاده شوند (بالاخص الگوریتم‌های درخت تصمیم، جنگل تصادفی، ماشین بردار پشتیبان، شبکه عصبی مصنوعی و ...). این موضوع در رابطه با الگوی اقتصادسنجی فضا-حالت نیز صادق است و این الگو توانایی مدل‌سازی و پیش‌بینی متغیرهای غیر خطی، نامانا و غیر نرمال را دارد. به نحوی که مدل‌های فضا حالت نیازی به فرض توزیع خاص برای داده‌های ورودی ندارند و می‌توانند به طور مستقیم داده‌های غیر نرمال را مدل کنند. همچنین با استفاده از فیلتر کالمن یا نسخه‌های تعمیم‌یافته آن (مانند فیلتر کالمن غیر خطی یا ذرات)، می‌توان تخمین‌های پویا از وضعیت سیستم محاسبه کرد. در مطالعه (Thayyib et al., 2023) نیز محققین به پیش‌بینی درآمدهای مالیاتی هند با استفاده از شبکه عصبی مصنوعی پرداخته‌اند، به نحوی که با وجود اینکه درآمدهای مالیاتی هند، غیر خطی، نامانا و غیر نرمال بوده است، اما متغیر درآمدهای مالیاتی را بدون هیچ پیش‌پردازشی، مدل‌سازی و پیش‌بینی کرده‌اند.

جدول ۱. آماره‌های توصیفی درآمدهای مالیاتی استان آذربایجان غربی

نام متغیر	نماد متغیر	میانگین (میلیون ریال)	میانه (میلیون ریال)	انحراف معیار	آزمون جبارک برا	
					مقدار	احتمال مربوطه
درآمدهای مالیاتی استان آذربایجان غربی	tax	۹۶۸،۸۰۸	۳۹۷،۶۰۷	۱،۴۸۸،۲۲۲	۵۱۳	۰.۰۰۰

منبع: نتایج تحقیق.

نتایج تجربی تحقیق

مقایسه عملکرد الگوهای یادگیری ماشین در پیش‌بینی درون نمونه‌ای و برون نمونه‌ای
درآمدهای مالیاتی استان آذربایجان غربی

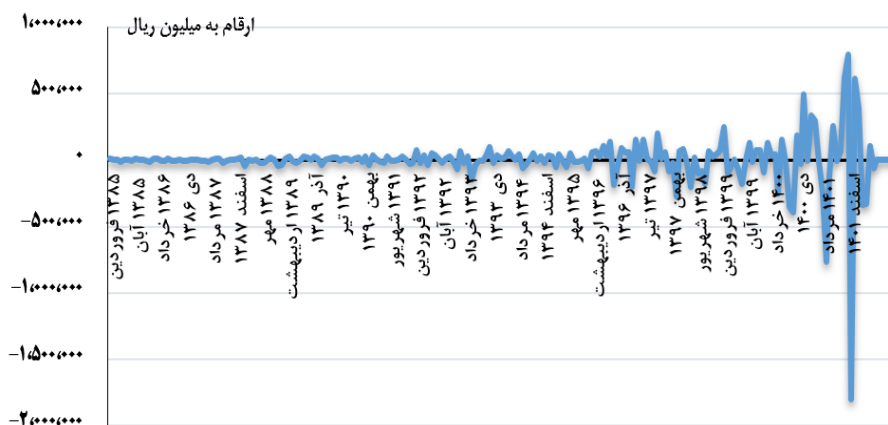
پیش‌بینی درون نمونه‌ای

درختان تصمیم‌گیری مستعد بیش‌برازش هستند. یک درخت به اندازه کافی عمیق می‌تواند هر مقدار هدف را در یک برگ مجزا کند، در این صورت پیش‌بینی‌های مدل دقیقاً با مقادیر هدف مشاهده‌شده در طول آموزش مطابقت دارند، اما احتمالاً تقریب خوبی برای داده‌های جدید دیده‌نشده که برای آموزش استفاده نشده‌اند ارائه نمی‌دهند. درختان تصمیم‌گیری همچنین نسبت به داده‌های ورودی بسیار حساس هستند، به طوری که تغییرات کوچک در داده‌های آموزشی می‌تواند به ساختارهای درختی کاملاً متفاوت منجر شود. جنگل‌های تصادفی این محدودیت‌ها را با ایجاد مجموعه‌ای از درختان تصمیم‌گیری که بر روی زیرمجموعه‌های تصادفی مختلف از داده‌های آموزشی (کیسه‌گذاری نمونه) با استفاده از زیرمجموعه‌های تصادفی مختلف از ویژگی‌ها (کیسه‌گذاری ویژگی) آموزش داده می‌شوند، برطرف می‌کنند. پیش‌بینی‌های جنگل تصادفی با میانگین‌گیری از پیش‌بینی‌های فردی درختان در مجموعه به‌دست می‌آیند. مکانیسم‌های کیسه‌گذاری نمونه و کیسه‌گذاری ویژگی همبستگی بین پیش‌بینی‌های درختان مختلف را کاهش می‌دهند، که مجموعه کلی را مقاوم‌تر و کمتر مستعد بیش‌برازش می‌کنند (الگو در محیط نرم افزاری R کدنویسی و اجرا شده است).

جهت پیش‌بینی کل درآمدهای مالیاتی استان آذربایجان غربی از رگرسیون تک متغیره پویا بر اساس وقفه بهینه در الگو استفاده می‌شود. در این روش کل درآمدهای مالیاتی بر اساس وقفه‌های خود مدل‌سازی شده و بر اساس روش جنگل تصادفی پیش‌بینی درون نمونه‌ای و برون نمونه‌ای انجام می‌گیرد.

مدل‌سازی درآمدهای مالیاتی استان آذربایجان غربی بر اساس ۲۰ درخت و ۱۰ وقفه انجام گرفت که بر اساس نمودار ۱، مقدار اختلاف برازش و پیش‌بینی درون نمونه‌ای کل درآمدهای مالیاتی استان آذربایجان غربی قابل مشاهده است.

نمودار ۱. اختلاف پیش‌بینی درون نمونه‌ای جنگل تصادفی از مقدار وصولی درآمدهای مالیاتی استان آذربایجان غربی

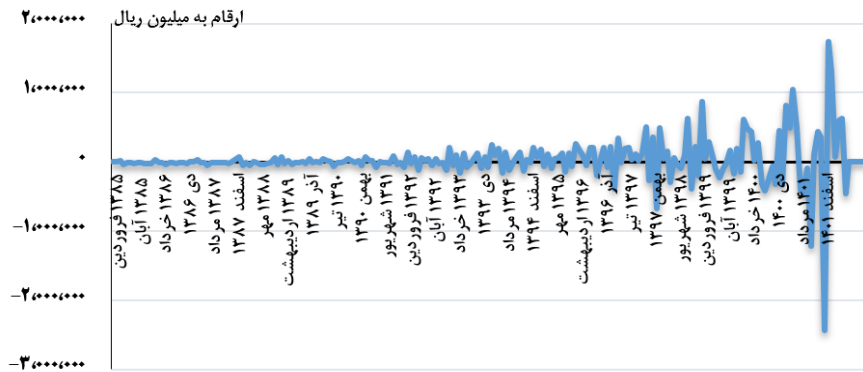


ماخذ: نتایج تحقیق.

رگرسیون بردار پشتیبانی (SVR) یک تکنیک یادگیری نظارت‌شده در ماشین‌های بردار پشتیبانی است که هدف آن یافتن ابرصفحه‌ای در فضای ویژگی‌های با ابعاد بالا است که بهترین تطابق را با داده‌های آموزشی دارد و خطای پیش‌بینی را برای وظایف رگرسیون به حداقل می‌رساند. رگرسیون بردار پشتیبانی یک تکنیک برای پیش‌بینی مقادیر پیوسته است. در پیش‌بینی سری‌های زمانی با SVR، این تکنیک به عنوان یک وظیفه رگرسیون در نظر گرفته می‌شود. الگوی رگرسیون بردار پشتیبانی بر اساس الگوی ARIMA بهینه با ۹ وقفه بهینه، با لحاظ توابع کرنل خطی و در سطح معناداری ۵ درصد برای داده‌های ماهانه کل درآمدهای مالیاتی استان آذربایجان غربی برآورد شد که در نمودار شماره ۲، اختلاف برازش و پیش‌بینی درون نمونه‌ای کل درآمدهای مالیاتی استان آذربایجان غربی قابل مشاهده است. (الگو در محیط نرم افزاری R کدنویسی و اجرا شده است).

با مقایسه نمودار شماره ۱ و ۲، به صورت گرافیکی و نموداری تقریباً مشخص است که اختلاف مقدار پیش‌بینی درون نمونه‌ای جنگل تصادفی با مقادیر واقعی درآمدهای مالیاتی نسبت به اختلاف درون نمونه‌ای روش رگرسیون بردار پشتیبانی با مقادیر واقعی مالیات کمتر است، البته در ادامه بر اساس معیارهای ارزیابی عملکرد موضوع دقیق‌تر بررسی خواهد شد.

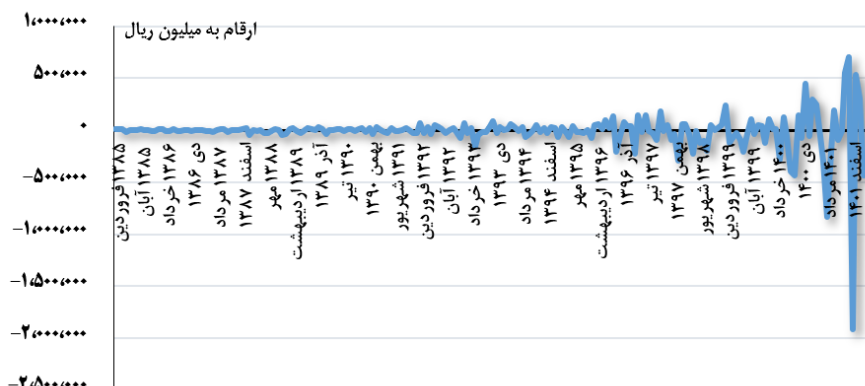
نمودار ۲. اختلاف پیش‌بینی درون نمونه‌ای رگرسیون بردار پشتیبان از مقدار درآمدهای وصولی مالیاتی استان آذربایجان غربی



ماخذ: نتایج تحقیق.

مدل LASSO یکی از الگوریتم‌های رگرسیون خطی است که برای پیش‌بینی سری‌های زمانی استفاده می‌شود. LASSO نوعی رگرسیون منظم‌سازی است که علاوه بر کمینه کردن خطاهای پیش‌بینی، سعی می‌کند برخی از ضرایب ویژگی‌ها را به صفر نزدیک کند. این کار به منظور انتخاب ویژگی‌ها و جلوگیری از بیش‌برازش انجام می‌شود. LASSO یک تکنیک رگرسیون است که از یک جریمه برای کنترل پیچیدگی مدل استفاده می‌کند. داده‌های سری زمانی به فرمتی که برای forecastML و glmnet مناسب است، تبدیل می‌شوند (جهت کار و برآورد در محیط نرم‌افزار R). با توجه به توضیحات داده شده، الگوریتم یادگیری استاندارد ماشین پیکربندی شد و پیش‌بینی درون نمونه‌ای برای داده‌های ماهانه درآمدهای مالیاتی استان آذربایجان غربی انجام گرفت که در نمودار شماره ۳ مقدار اختلاف پیش‌بینی و برازش شده با مقدار واقعی ترسیم شده است.

نمودار ۳. اختلاف پیش‌بینی درون نمونه‌ای یادگیری استاندارد ماشین از مقدار وصولی درآمدهای مالیاتی استان آذربایجان غربی



ماخذ: نتایج تحقیق.

جهت پیش‌بینی درآمدهای کل مالیاتی استان آذربایجان غربی با الگوریتم یادگیری عمیق ماشین مبتنی بر منطق فازی بایستی برخی تنظیمات و پارامترها ابتدا تنظیم شود که در ادامه توضیح داده می‌شود (الگو در محیط نرم‌افزاری MATLAB کدنویسی و اجرا شده است).

۱. روش ایجاد: این قسمت از کد برای ایجاد مدل‌های شبکه عصبی مصنوعی با استفاده از مدل غیر خطی ARX استفاده می‌شود.

۲. **nstep**: تعداد گام‌های زمانی برای مقایسه خروجی مدل با داده‌های واقعی. (۱۰۰)

۳. **K**: تعداد ماه‌هایی که می‌خواهید برای آینده پیش‌بینی کنید. (تعداد ۹ ماه)

۴. **Delays**: این آرایه مشخص می‌کند که چه تاخیرهایی در داده‌های سری زمانی استفاده خواهد شد. به عبارتی، این مقادیر نشان می‌دهند که برای هر نمونه ورودی، چه تعداد گام زمانی گذشته به عنوان ورودی به مدل داده می‌شود. (تعداد ۹ انتخاب شد)

۵. **Time Series Function**: این تابع داده‌های سری زمانی را با توجه به تاخیرهای مشخص شده ایجاد می‌کند. خروجی‌های آن ورودی‌ها (Inputs) و اهداف (Targets) هستند که به ترتیب داده‌های ورودی و خروجی برای آموزش شبکه عصبی را تشکیل می‌دهند.

۶. **Shuffling Data**: این بخش از کد داده‌ها را به صورت تصادفی مخلوط می‌کند تا ترتیب داده‌ها در مراحل آموزش و آزمون به طور یکنواخت تقسیم شود.

۷. **Train/Test Split**: در این بخش، داده‌ها به دو مجموعه آموزشی و آزمون تقسیم می‌شوند. p Train درصد داده‌های آموزشی را مشخص می‌کند که در اینجا ۸۰٪ تعیین شده است. $TrainInd$ شاخص‌های داده‌های آموزشی و $TrainInputs$ و $TrainTargets$ ورودی‌ها و اهداف آموزشی را نشان می‌دهند.

۸. **FCMFIS Generation Parameters**: این بخش مربوط به تنظیم پارامترهای روش تولید سیستم استنتاج فازی (Fuzzy Inference System) با استفاده از الگوریتم خوشه‌بندی فازی (FCM) است:

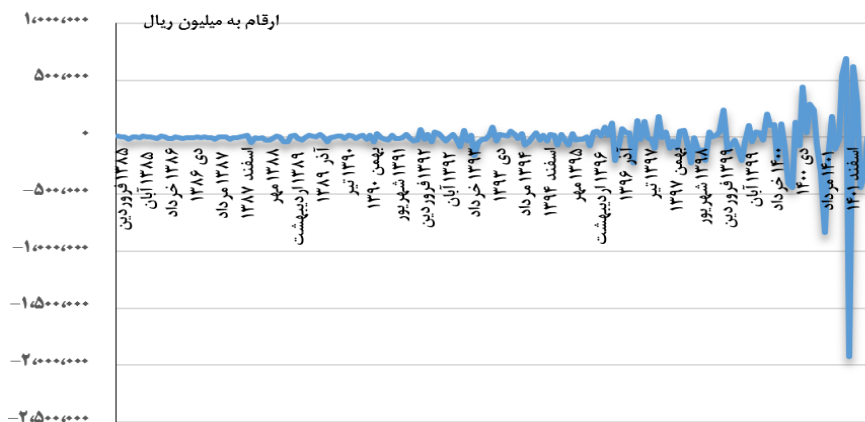
nCluster: تعداد خوشه‌ها (**Clusters**): که داده‌ها به آن‌ها تقسیم می‌شوند، در اینجا ۲۰ خوشه تعیین شده است.

Exponent: مقدار نمایی که در محاسبات FCM استفاده می‌شود. این مقدار میزان فازی (Fuzziness) را در خوشه‌بندی تعیین می‌کند. مقدار بالاتر مدل را بیشتر انعطاف‌پذیر می‌کند.

MaxIt: حداکثر تعداد تکرارهای الگوریتم FCM در اینجا ۵۰۰ تکرار تعیین شده است. **MinImprovement**: حداقل بهبود مورد نیاز برای توقف الگوریتم. اگر بهبود کمتر از این مقدار باشد، الگوریتم متوقف می‌شود. این مقدار در اینجا ۱۰-۵ تعیین شده است. **DisplayInfo**: اگر این مقدار برابر ۱ باشد، اطلاعات مربوط به پیشرفت الگوریتم در هر تکرار نمایش داده می‌شود.

با توجه به توضیحات داده شده، شبکه عصبی مصنوعی با منطق فازی پیکربندی شد و پیش‌بینی درون نمونه‌ای برای داده‌های ماهانه درآمدهای مالیاتی استان آذربایجان غربی انجام گرفت که در نمودار شماره ۴ مقدار اختلاف پیش‌بینی و برازش شده با مقدار واقعی درآمدهای مالیاتی ترسیم شده است. مطابق با نمودار شماره ۴، روش شبکه عصبی مصنوعی با منطق فازی نیز عملکرد خوبی در پیش‌بینی درآمدهای مالیاتی استان آذربایجان غربی دارد، البته جهت مقایسه عملکرد آن با سه روش قبلی در ادامه از معیارهای ارزیابی عملکرد استفاده خواهد شد.

نمودار ۴. اختلاف پیش‌بینی درون نمونه‌ای یادگیری عمیق ماشین از مقدار وصولی درآمدهای مالیاتی استان آذربایجان غربی



ماخذ: نتایج تحقیق.

مطابق با جدول ۲، بر اساس معیار RMSE یادگیری عمیق ماشین بهترین عملکرد در پیش‌بینی درون نمونه‌ای را داشته است (البته عملکرد یادگیری استاندارد یادگیری ماشین نیز اختلاف بسیار کمی با یادگیری عمیق ماشین دارد). همچنین بر اساس معیار MAPE کمترین درصد خطا (۹.۸۸) مربوط به یادگیری عمیق ماشین است. به این معنا است که به طور میانگین، پیش‌بینی‌های یادگیری عمیق ماشین حدود ۹.۸۸ درصد از مقادیر واقعی خطا دارند. به عبارت دیگر، خطای نسبی یادگیری عمیق ماشین در پیش‌بینی‌ها، به طور میانگین ۹.۸۸ درصد است. MAPE به صورت درصد بیان می‌شود و نشان می‌دهد که مدل چقدر خوب توانسته است مقادیر واقعی را پیش‌بینی کند. اگر MAPE برابر با ۹.۸۸ باشد، این عدد به این معنی است که مدل دقت نسبتاً خوبی دارد، زیرا این میزان خطا معمولاً به عنوان یک خطای نسبی پایین در بسیاری از کاربردها در نظر گرفته می‌شود.

جدول ۲. مقایسه قدرت پیش‌بینی دورن‌نمونه‌ای ۴ الگوی یادگیری ماشین

الگو	معیار RMSE	معیار MAPE
جنگل تصادفی	۱۸۶،۴۸۶	۱۰.۱۳ درصد
رگرسیون بردار پشتیبان	۳۲۳،۶۱۹	۲۳.۴۶ درصد
یادگیری استاندارد ماشین	۱۷۱،۲۴۵	۹.۹۰ درصد
یادگیری عمیق ماشین	۱۵۰،۲۴۵	۹.۸۸ درصد

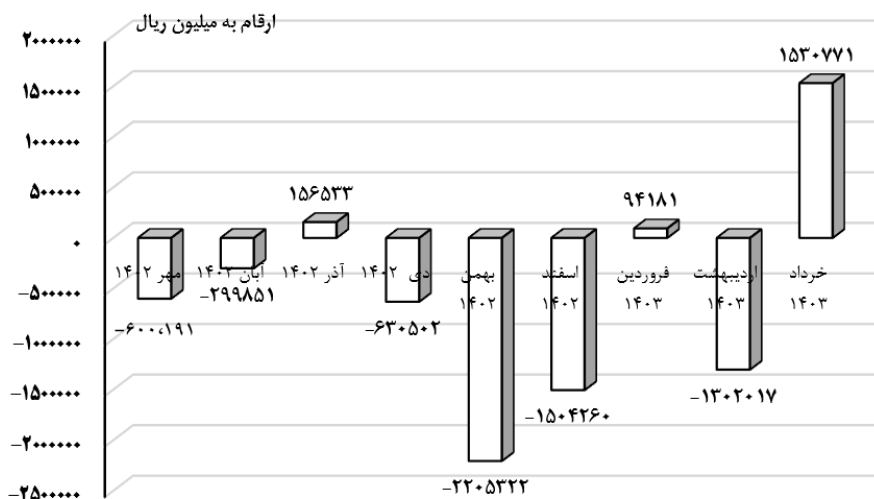
ماخذ: نتایج تحقیق.

پیش‌بینی برون نمونه‌ای الگوریتم‌های یادگیری ماشین

در نمودارهای ۵ الی ۸ روند پیش‌بینی برون نمونه‌ای ۴ الگوریتم مختلف یادگیری ماشین نشان داده شده است. پیش‌بینی که برای ماه‌های مهر ۱۴۰۲ الی خرداد ۱۴۰۳ انجام شده است، می‌توان استنباط کرد که عملکرد یادگیری ماشین استاندارد و یادگیری عمیق ماشین بسیار بهتر از دو الگوریتم دیگر است و به مقدار واقعی داده‌ها بسیار نزدیک است.

مطابق با نمودار ۵ که اختلاف پیش‌بینی برون نمونه‌ای روش جنگل تصادفی با مقادیر واقعی درآمدهای مالیاتی را بین مهر ۱۴۰۲ الی خرداد ۱۴۰۳ نشان می‌دهد، قابل استنباط است که در برخی از ماه‌ها بیشتر از مقدار درآمد واقعی مالیاتی پیش‌بینی کرده (مقادیر مثبت در نمودار) و در برخی از ماه‌ها نیز مقدار پیش‌بینی کمتر از مقدار واقعی بوده است (به طور مثال در مهر سال ۱۴۰۲، به مقدار ۶۰۰،۱۹۱ میلیون ریال، از مقدار واقعی وصولی کمتر پیش‌بینی کرده است) در بهمن و اسفند سال ۱۴۰۲ و خرداد سال ۱۴۰۳، پیش‌بینی الگوی جنگل تصادفی اختلاف قابل توجهی با مقادیر واقعی دارد.

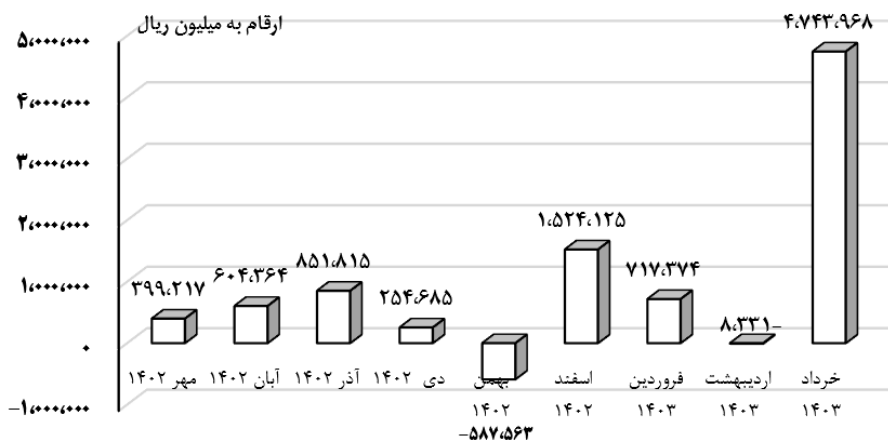
نمودار ۵. اختلاف پیش‌بینی برون نمونه‌ای جنگل تصادفی از مقدار درآمدهای مالیاتی استان آذربایجان غربی



ماخذ: نتایج تحقیق.

مطابق با نمودار ۶ که اختلاف پیش‌بینی برون نمونه‌ای روش رگرسیون بردار پشتیبان با مقادیر واقعی درآمدهای مالیاتی را بین مهر ۱۴۰۲ الی خرداد ۱۴۰۳ نشان می‌دهد، قابل استنباط است که در بسیاری از ماه‌ها بیشتر از مقدار درآمد واقعی مالیاتی پیش‌بینی کرده است. به نحوی که در اسفند سال ۱۴۰۲ و خرداد سال ۱۴۰۳، میزان این اختلاف به حداکثر مقدار خود رسیده است.

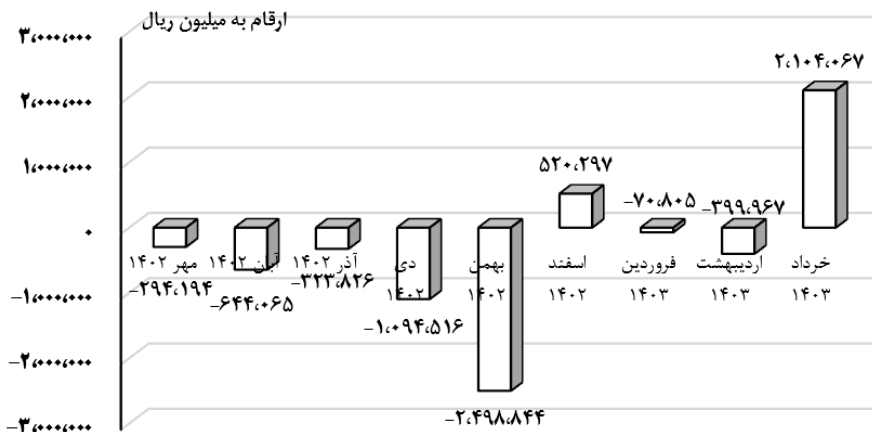
نمودار ۶. اختلاف پیش‌بینی برون نمونه رگرسیون بردار پشتیبان از مقدار درآمدهای مالیاتی استان



ماخذ: نتایج تحقیق.

مطابق با نمودار ۷ که اختلاف پیش‌بینی برون نمونه‌ای روش یادگیری استاندارد ماشین با مقادیر واقعی درآمدهای مالیاتی را بین مهر ۱۴۰۲ الی خرداد ۱۴۰۳ نشان می‌دهد، قابل استنباط است که در برخی از ماه‌ها بیشتر و در برخی از ماه‌ها کمتر از مقدار درآمد واقعی مالیاتی پیش‌بینی کرده است. به نحوی که در بهمن سال ۱۴۰۲ به مقدار قابل توجهی کمتر از مقادیر واقعی و در خرداد سال ۱۴۰۳ به میزان قابل توجهی بیشتر از مقدار واقعی پیش‌بینی کرده است.

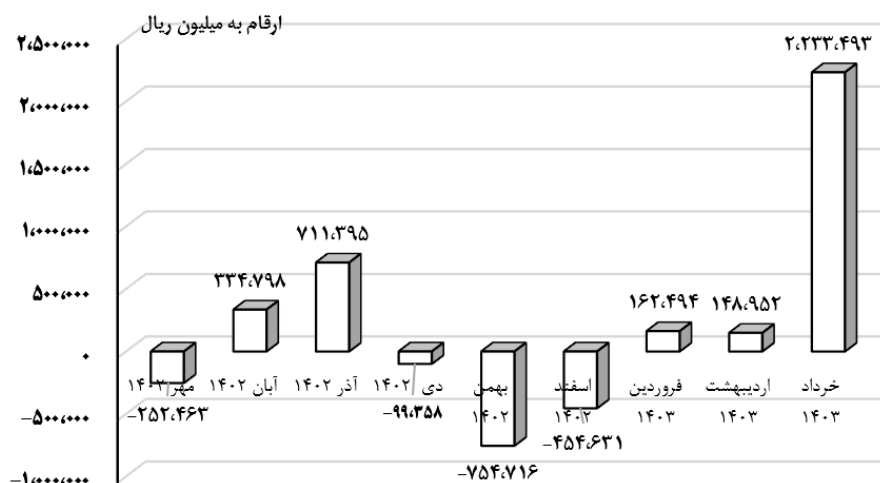
نمودار ۷. اختلاف پیش‌بینی برون نمونه‌ای یادگیری استاندارد ماشین از مقدار درآمدهای مالیاتی استان آذربایجان غربی



ماخذ: نتایج تحقیق.

مطابق با نمودار ۸ که اختلاف پیش‌بینی برون نمونه‌ای روش یادگیری عمیق ماشین منطق فازی با مقادیر واقعی درآمدهای مالیاتی را بین مهر ۱۴۰۲ الی خرداد ۱۴۰۳ نشان می‌دهد، قابل استنباط است که در برخی از ماه‌ها بیشتر و در برخی از ماه‌ها کمتر از مقدار درآمد واقعی مالیاتی پیش‌بینی کرده است. البته با مقایسه این نمودار با نتایج حاصل شده، می‌توان استنباط کرد که عملکرد الگوریتم یادگیری عمیق بهتر از سایر روش‌های قبلی بوده است و به جز خرداد سال ۱۴۰۳، در بقیه ماه‌ها اختلاف کمتری با مقادیر واقعی دارد.

نمودار ۸. اختلاف پیش‌بینی برون نمونه‌ای یادگیری عمیق ماشین از مقدار درآمدهای مالیاتی استان آذربایجان غربی



ماخذ: نتایج تحقیق.

حال چنانچه عملکرد پیش‌بینی الگوهای مذکور به صورت آماری با یکدیگر مقایسه شوند، این مقایسه به صورت جدول ۳ خواهد بود.

با توجه به عملکرد الگوهای مختلف یادگیری ماشین بر اساس هر دو معیار RMSE و MAPE یادگیری عمیق ماشین در پیش‌بینی برون نمونه‌ای عملکرد بهتر و قوی‌تری نسبت به سه روش دیگر در پیش‌بینی درآمدهای مالیاتی استان آذربایجان غربی دارد.

جدول ۳. مقایسه قدرت پیش‌بینی دورن نمونه‌ای ۴ الگوی یادگیری ماشین

الگو	معیار RMSE	معیار MAPE
جنگل تصادفی	۱,۱۵۶,۹۰۱	۱۵.۵۸ درصد
رگرسیون بردار پشتیبان	۱,۷۳۲,۱۵۳	۲۱.۵۵ درصد
یادگیری استاندارد ماشین	۱,۱۹۷,۷۳۹	۱۵.۶۴ درصد
یادگیری عمیق ماشین	۸۵۰,۱۷۳	۱۱.۲۱ درصد

ماخذ: نتایج تحقیق.

مقایسه عملکرد الگوی فضا حالت با الگوریتم یادگیری عمیق ماشین در پیش‌بینی درون نمونه‌ای و برون نمونه‌ای درآمدهای مالیاتی استان آذربایجان غربی

الگوی فضا حالت دارای یک معادله خودرگرسیون با ۲ وقفه و قسمت الگوی حالت با وضعیت تصادفی روند، شیب، پارامتر فصل و چرخه است. نتایج بهترین الگو از میان الگوهای متعدد فضا حالت (الگوی ساختاری با متغیرهای غیر قابل مشاهده) در جدول (۴) آمده است. آنچه بیشتر از تفسیر نتایج جدول زیر اهمیت دارد، پیش‌بینی برون نمونه‌ای و معیارهای پیش‌بینی حاصل از تخمین الگو است که می‌تواند نشان دهد الگوی تخمینی از سایر الگوهای رقیب (یادگیری ماشین) در پیش‌بینی درآمدهای مالیاتی عملکرد بهتری داشته یا نه؟ که در قسمت بعدی به این موضوع پرداخته می‌شود.

جدول ۴. الگوی فضا حالت درآمدهای مالیاتی استان آذربایجان غربی

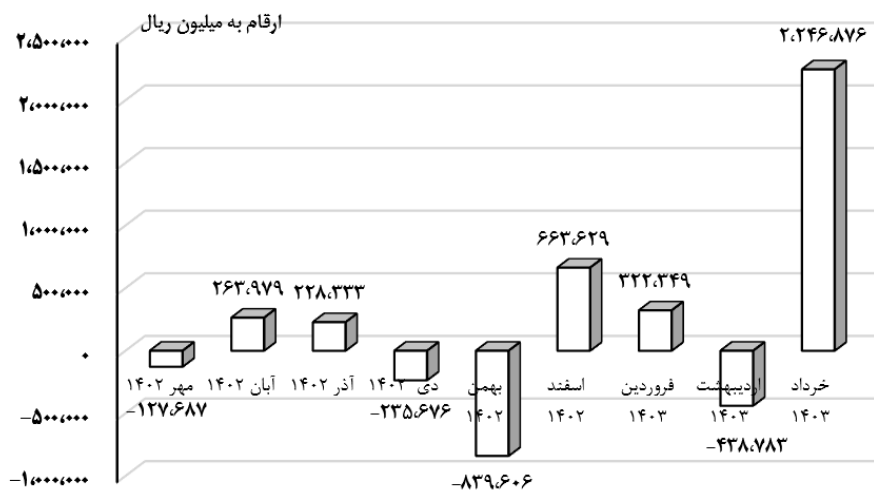
معناداری	ضریب	نام متغیر یا پارامتر
قسمت خودرگرسیون		
۰.۰۰۰	۰.۰۰۰	AR(1)
۶۶۲.۰	۱۰۶۲.۱+E	AR(2)
قسمت حالت الگو		
۰.۰۰۰	۵۲۳۳۷۸۱	LEVEL سطح روند
۰.۰۰۰	۱۴۶۵۶۹	SLOPE شیب روند
اوپر پارامترها		
۰.۰۰۰	۰.۰۰۰	LEVEL واریانس سطح روند
۰.۰۰۰	۰۷۳۴.۵+E	SLOPE واریانس شیب روند
۰۰۹.۰	۰۸۶۰.۲+E	Seasonal واریانس پارامتر فصل (ماه)
۵۰۷.۰	۱۰۲۴.۱+E	Cycle واریانس پارامتر چرخه
۱	۱۰۴۴.۲+E	Irregular واریانس پارامتر جزء نامنظم

ماخذ: نتایج تحقیق.

اختلاف پیش‌بینی برون نمونه‌ای حاصل از مدل‌سازی الگوی فضا حالت با مقادیر واقعی درآمدهای مالیاتی استان آذربایجان غربی در نمودار ۹ آورده شده است. مطابق با روند اختلاف پیش‌بینی با درآمدهای وصولی مالیاتی استان آذربایجان غربی از مهر سال ۱۴۰۲ الی خرداد ۱۴۰۳

می‌توان استنباط کرد که الگوی مذکور تا حد قابل قبولی به مقادیر واقعی نزدیک است و صرفاً در خرداد ۱۴۰۳ اختلاف قابل توجهی بین مقدار پیش‌بینی و مقدار واقعی درآمدهای مالیاتی استان آذربایجان غربی وجود دارد.

نمودار ۹. اختلاف پیش‌بینی برون نمونه‌ای الگوی فضا حالت از مقدار درآمدهای مالیاتی استان آذربایجان غربی



ماخذ: نتایج تحقیق.

چنانچه نمودار ۹ به صورت آماری و بر اساس معیارهای عملکرد پیش‌بینی خلاصه شود، جدول ۵ به دست می‌آید.

در قسمت قبلی نشان داده شد که در بین الگوریتم‌های جنگل تصادفی، رگرسیون بردار پشتیبان، یادگیری ماشین مبتنی بر رگرسیون لاسو و یادگیری عمیق ماشین، الگوی یادگیری عمیق ماشین بهترین عملکرد را در پیش‌بینی درون نمونه‌ای و برون نمونه‌ای از درآمدهای مالیاتی استان آذربایجان غربی دارد. با این حال الگوی فضا حالت نیز به عنوان یک ابزار قدرتمند در مدل‌سازی و پیش‌بینی متغیرهای مالی و اقتصادی است، به خصوص زمانی که روندهای فصلی، ماهانه و یا روزانه در آن متغیر وجود داشته باشد. همچنین وجود اجزاء غیر قابل مشاهده، اجزاء نامنظم و وجود شیب تصادفی در مدل‌سازی الگوهای مالی و اقتصادی از ویژگی‌های بارز این الگو است.

نتایج جدول ۵ حاکی از این موضوع است که با توجه به عملکرد الگوهای یادگیری عمیق ماشین و الگوی فضا حالت، بر اساس هر دو معیار $RMSE$ و $MAPE$ الگوی فضا حالت در پیش‌بینی برون نمونه‌ای عملکرد بهتر و قوی‌تری نسبت به روش یادگیری عمیق ماشین در پیش‌بینی درآمدهای مالیاتی استان آذربایجان غربی دارد. به نحوی که بر اساس معیار $MAPE$ خطای پیش‌بینی الگوی فضا حالت و یادگیری عمیق ماشین به ترتیب برابر با ۱۱.۰۵ و ۱۱.۲۱ درصد است.

جدول ۵. عملکرد الگوی فضا-حالت در مقایسه با یادگیری عمیق ماشین در پیش‌بینی برون نمونه‌ای درآمدهای مالیاتی استان آذربایجان غربی

الگو	معیار $RMSE$	معیار $MAPE$
فضا-حالت	۸۴۲،۴۸۷	۱۱.۰۵ درصد
یادگیری عمیق ماشین	۸۵۰،۱۷۳	۱۱.۲۱ درصد

ماخذ: نتایج تحقیق.

جمع‌بندی و نتیجه‌گیری

نتایج نشان داد بر اساس معیار $RMSE$ یادگیری عمیق ماشین بهترین عملکرد در پیش‌بینی درون نمونه‌ای را در بین الگوریتم‌های یادگیری ماشین داشته است (البته عملکرد یادگیری استاندارد یادگیری ماشین نیز اختلاف بسیار کمی با یادگیری عمیق ماشین دارد). همچنین بر اساس معیار $MAPE$ کمترین درصد خطا (۹.۸۸ درصد) مربوط به یادگیری عمیق ماشین است. به این معنا که خطای نسبی الگوریتم یادگیری عمیق ماشین در پیش‌بینی‌ها، به طور میانگین ۹.۸۸ درصد است. در بین الگوریتم‌های جنگل تصادفی، رگرسیون بردار پشتیبان، یادگیری ماشین مبتنی بر رگرسیون لاسو و یادگیری عمیق ماشین، الگوی یادگیری عمیق ماشین بهترین عملکرد را در پیش‌بینی درون نمونه‌ای و برون نمونه‌ای از درآمدهای مالیاتی استان آذربایجان غربی دارد. با این حال الگوی فضا حالت نیز به عنوان یک ابزار قدرتمند در مدل‌سازی و پیش‌بینی متغیرهای مالی و اقتصادی است، به خصوص زمانی که روندهای فصلی، ماهانه و یا روزانه در آن متغیر وجود داشته باشد. همچنین وجود اجزاء غیر قابل مشاهده، اجزاء نامنظم و وجود شیب تصادفی در مدل‌سازی الگوهای مالی و اقتصادی از ویژگی‌های بارز این الگو است.

نتایج حاکی از این موضوع است که با توجه به عملکرد الگوهای یادگیری عمیق ماشین و الگوی فضا حالت، بر اساس هر دو معیار RMSE و MAPE الگوی فضا حالت در پیش‌بینی برون نمونه‌ای عملکرد بهتر و قوی‌تری نسبت به روش یادگیری عمیق ماشین در پیش‌بینی درآمدهای مالیاتی استان آذربایجان غربی دارد. به نحوی که بر اساس معیار MAPE خطای پیش‌بینی الگوی فضا حالت و یادگیری عمیق ماشین به ترتیب برابر با ۱۱.۰۵ و ۱۱.۲۱ درصد است.

بر این اساس پیشنهادهای زیر برای سیاست‌گذاران کلان کشور، بالاخص سیاست‌گذاران مالی و مالیاتی ارائه می‌شود.

۱. با توجه به اهمیت پیش‌بینی برای برنامه‌ریزی و مدیریت کلان کشور پیشنهاد می‌شود، در پیش‌بینی جهت تعیین نقشه راه درآمدهای مالیاتی کشور و تعیین سهم هریک از استان‌ها در میزان وصولی کل مالیات‌های کشور، از ابزارهای یادگیری ماشین و الگوهای فضا حالت استفاده شود.

۲. مدل‌های یادگیری عمیق و الگوهای پیشرفته اقتصادسنجی می‌توانند به پیش‌بینی درآمد و مخارج دولت پردازند و از این جهت می‌توان با توجه به نقش آن‌ها در رشد اقتصادی، میزان رشد اقتصادی منتج از تغییرات این سیاست‌ها را پیش‌بینی کرد، لذا پیشنهاد می‌شود از این ابزار جهت تعیین رشد اقتصادی نیز استفاده شود.

۳. پیشنهاد می‌شود سیاست‌گذاران کلان کشور جهت مدیریت صحیح، علمی و منطقی مخارج و درآمدهای دولت و جهت شفاف‌سازی در بین آحاد مردم، متخصصین و دانشگاهیان، هر ساله با توسل به متخصصین کشور، از پیش‌بینی‌های علمی منتج از روش‌های اقتصادسنجی، آماری، ریاضی و یادگیری‌های ماشین این متخصصین، درآمدهای مالیاتی و مخارج سالانه هر استان را بر اساس این روش‌ها تعیین نمایند تا پشوانه علمی و اصولی و منطقی در افزایش و کاهش درآمدهای مالیاتی کشور و استان‌ها وجود داشته باشد.

تعارض منافع

تعارض منافع وجود ندارد.

سپاسگزاری

از تمامی افرادی که برای انجام هرچه بهتر این پژوهش، یاری رساندند، تشکر می‌شود.

ORCID

Rahim Dabbagh  <http://orcid.org/0000-0003-2888-3928>

Ahmad Ezzati Shourgoli  <http://orcid.org/0000-0001-6632-5893>

Raheleh Shabanpoor  <https://orcid.org/0009-0002-7946-3050>

منابع

۱. دباغ، رحیم، فرزانه، محمد و محسنی، بهزاد. (۱۴۰۴). شناسایی چالش‌های فناوری‌های صنعت ۴/۰ برای عملیات پایدار در صنایع کوچک و متوسط. پژوهش در مدیریت تولید و عملیات، ۱۶(۱)، ۲۴-۱.
۲. دباغ، رحیم و گلمرادی، باقری (۱۳۹۸). بررسی عملکرد مالی بانک‌های اسلامی و غیراسلامی در کشورهای منتخب با استفاده از مدل CAMEL. فصل‌نامه مطالعاتی در مدیریت بانکی و بانکداری اسلامی، ۴(پاییز)، ۸۵-۱۱۴.
۳. زارعی، پریسا، جلالی، سید عبدالمجید و صادقی، زین العابدین (۱۳۹۸). شبیه‌سازی و پیش‌بینی اثر مالیات سبز بر مصرف و شدت انرژی در ایران با استفاده از الگوریتم ژنتیک. پژوهشنامه مالیات، ۲۷(۴۲)، ۱۰۳-۱۲۵.
۴. سعادت مهر، مسعود. (۱۴۰۰). بررسی تأثیر متغیرهای کلان اقتصاد بر درآمدهای مالیاتی دولت در ایران. دوفصلنامه جستارهای اقتصادی ایران با رویکرد اقتصاد اسلامی، ۱۸(۳۵)، ۱۶۹-۱۹۱.
۵. سلامی، حبیب اله و مافی، حسن. (۱۳۹۷). پیش‌بینی قیمت صادراتی پسته ایران مبتنی بر چرخه‌های تجاری: کاربست الگوی سری زمانی ساختاری. تحقیقات اقتصاد و توسعه کشاورزی ایران، ۴۹(۴)، ۵۷۱-۵۵۹.
۶. شیرانی فخر، زهره، خوش اخلاق، رحمان، و شریفی، علی مراد. (۱۳۹۳). تخمین تابع تقاضای گاز طبیعی بخش صنعت ایران با استفاده از مدل سری زمانی ساختاری. مطالعات اقتصادی کاربردی ایران (مطالعات اقتصادی کاربردی)، ۳(۱۱)، ۱۲۹-۱۵۷.
۷. عباسیان، عزت‌الله، محمودی، وحید و شاکر، ایمان. (۱۳۹۰). تحلیل خطای پیش‌بینی درآمدهای مالیاتی دولت در اقتصاد ایران. تحقیقات مالی، ۱۳(۳۲)، ۱۰۹-۱۳۲.
۸. عزتی شورگلی، احمد، زینال‌زاده شیخ سرمست، مهتاب و عیسوی، هیرو. (۱۴۰۲). پیش‌بینی اثر وصل شدن تراکنش دستگاه کارت‌خوان‌ها به سیستم سازمان امور مالیاتی در افزایش وصولی سازمان امور مالیاتی استان آذربایجان غربی. پژوهشنامه مالیات، ۳۴(۶۰)، ۱۳۴-۱۶۵.
۹. نصراللهی، زهرا و شاکر اردکانی، ایمان. (۱۳۹۵). بررسی عملکرد طراحان بودجه در پیش‌بینی درآمدهای دولت در اقتصاد ایران. پژوهش‌های اقتصادی (رشد و توسعه پایدار)، ۱۶(۴)، ۱۲۷-۱۴۷.

References

1. Abbasian, E., Mahmoudi, V., & Shaker, I. (2012). Forecast Error Analysis of State Tax Revenues in Iran. *Financial Research Journal*, 13(32), 109-132. [In Persian]
2. Abdi, M., Hamidialamdari, S., & Pourhassanamiri, M. (2011). An Evaluation of Forecasting Methods and Optimal Combination Models to Predict Tax Revenues. *Journal of Tax Research*, 19(11), 85-120. [In Persian]
3. Altaf, S., Khan, S. Y., Ali, M. W., & Khan, M. W. (2021). A mathematical forecasting model to estimate tax to GDP ratio. *International Journal of Advanced Research in Engineering and Technology*, 12(3), 291-297.
4. Andrejovska, A., & Pulikova, V. (2018). Tax revenues in the context of economic determinants. *Montenegrin Journal of Economics*, 14(1), 133-141.
5. Arfaee, M., Bahari, A., & Khalilzadeh, M. (2022). A novel prediction model for educational planning of human resources with data mining approach: a national tax administration case study. *Education and Information Technologies*, 27(2), 2209-2239.
6. Battiston, P., Gamba, S., & Santoro, A. (2024). Machine learning and the optimization of prediction-based policies. *Technological Forecasting and Social Change*, 199, 123080.
7. Buettner, T., & Kauder, B. (2010). Revenue forecasting practices: Differences across countries and consequences for forecasting performance. *Fiscal Studies*, 31(3), 313-340.
8. Castro, G. Á., & Camarillo, D. B. R. (2014). Determinants of tax revenue in OECD countries over the period 2001–2011. *Contaduría Y Administración*, 59(3), 35-59.
9. Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Systems with Applications*, 112, 353-371.
10. Dabbagh, R., Golmoradi, H., & Bagri, A. (2019). Financial performance comparison of Islamic and conventional banking in selected countries using the CAMEL model. *Quarterly Studies in Banking Management and Islamic Banking*, 4 (Autumn), 85-114. [In Persian]
11. Dabbagh, R., Farzan, M. and Mohseni, B. (2025). Identifying the Challenges of Industry 4.0 Technologies for Sustainable Operations in SMEs. *Research in Production and Operations Management*, 16(1), 1-24. [In Persian]
12. Ezzati Shourghouli, A., Zeinalzadeh, M., & Isavi, H. (2024). Predicting the effect of connecting the transaction of card readers to the system of the tax affairs organization in increasing the collection of the tax affairs organization of West. *J Tax Res*, 31 (60), 133-165. [In Persian]

13. Febriminanto, R. D., & Wasesa, M. (2022). Machine Learning Analytics for Predicting Tax Revenue Potential. *Indonesian Treasury Review: Jurnal Perbendaharaan, Keuangan Negara dan Kebijakan Publik*, 7(3), 193-205.
14. Fikriya, A. A., & Hikmawati, S. (2020). Support Vector Machine Predictive Analysis Implementation: Case Study of Tax Revenue in Government of South Lampung. In *Proceeding International Conference on Science and Engineering* (Vol. 3, pp. 323-327).
15. Hajej, Z., Rezg, N., & Gharbi, A. (2017). Ecological optimization for forecasting production and maintenance problem based on carbon tax. *The International Journal of Advanced Manufacturing Technology*, 88, 1595-1606.
16. Holkova, B., Malichova, E., Falat, L., & Pancikova, L. (2023). Determinants of tax ethics in society: Statistical and logistic regression approach. *Mathematics*, 11(10), 2341.
17. Huang, W., Lai, K. K., Nakamori, Y., Wang, S., & Yu, L. (2007). Neural networks in finance and economics forecasting. *International Journal of Information Technology & Decision Making*, 6(01), 113-140.
18. Kalaš, B., Todorović, J. Đ., & Đorđević, M. (2020). Panel estimating effects of macroeconomic determinants on tax revenue level in European Union. *Industrija*, 48(3).
19. Karpowicz, A., & Majewska, E. (2018). Corporate Income Tax Revenue Determinants: How Important is the Tax Rate? *Economic and Social Development: Book of Proceedings*, 361-369.
20. Koirala, T.P. (2012) Government Revenue Forecasting in Nepal; *NRB Economic Review*, Vol. 24: 47-60.
21. Kuan, C. M., & White, H. (1994). Artificial neural networks: An econometric perspective. *Econometric Reviews*, 13(1), 1-91.
22. Lahiri, K., & Yang, C. (2022). Boosting tax revenues with mixed-frequency data in the aftermath of COVID-19, The case of New York. *International Journal of Forecasting*, 38(2), 545-566.
23. Lin, C. H., Lin, I. C., Wu, C. H., Yang, Y. C., & Roan, J. (2012). The application of decision tree and artificial neural network to income tax audit: the examples of profit-seeking enterprise income tax and individual income tax in Taiwan. *Journal of the Chinese Institute of Engineers*, 35(4), 401-411.
24. Li-Xia, L., Yi-Qi, Z., & Liu, X. Y. (2011). Tax forecasting theory and model based on SVM optimized by PSO. *Expert Systems with Applications*, 38(1), 116-120.
25. Ma, L., Shi, M., Li, Y., & Liu, Y. (2019). Prediction of Total Tax Value Based on Grey BP Correction Model. *Journal of Economics, Business and Management*, 7(2), 45-49.
26. Malikov, T. (2021). Methodological approaches to assessing and forecasting the tax potential of the region. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(11), 7056-7060.

27. Maulia, E., & Sofyan, H. (2018). Tax revenue and inflation rate predictions in Banda Aceh using Vector Error Correction Model (VECM). In IOP Conference Series: Materials Science and Engineering (Vol. 352, No. 1, p. 012056). IOP Publishing.
28. Milenković, N., Kalaš, B., Mirović, V., & Andrašić, J. (2020). The impact of macroeconomic determinants and tax form on inflation in selected Balkan countries. *Serbian Journal of Management*, 15(1), 7-18.
29. Molapo, M. A., Olaomi, J. O., & Ama, N. O. (2019). Bayesian vector autoregression method as an alternative technique for forecasting South African tax revenue. *Southern African Business Review*, 23(1).
30. Murorunkwere, B. F., Haughton, D., Nzabanita, J., Kipkoge, F., & Kabano, I. (2023). Predicting tax fraud using supervised machine learning approach. *African Journal of Science, Technology, Innovation and Development*, 15(6), 731-742.
31. Nasrollahi, Z., & Shaker Ardakani, I. (2016). Investigating the Performance of Budget Designers in Forecasting Government Revenues in the Iranian Economy. *QJER*, 16(4), 127-147. [In Persian]
32. Neog, Y., & Gaur, A. K. (2020). Macro-economic determinants of tax revenue in India: an application of dynamic simultaneous equation model. *International Journal of Economic Policy in Emerging Economies*, 13(1), 13-35.
33. Nosratabadi, S., Mosavi, A., Duan, P., Ghamisi, P., Filip, F., Band, S. S., & Gandomi, A. H. (2020). Data science in economics: comprehensive review of advanced machine learning and deep learning methods. *Mathematics*, 8(10), 1799.
34. Pradhan, R. P., Arvin, M. B., Nair, M., Bennett, S. E., & Bahmani, S. (2024). Some determinants and mechanics of economic growth in middle-income countries: The role of ICT infrastructure development, taxation and other macroeconomic variables. *The Singapore Economic Review*, 69(01), 297-333.
35. Ruhi, M., Abbasian, E., Momeni, M., & Amouzad, M., (2017). Prediction of corporate income tax collection in Iran, using Markov chain and discrete spectrum analysis. *J Tax Res*, 26(38), 107-129. [In Persian].
36. Saadat Mehr, M. (2021). Investigation of the impact of macroeconomic variables on government tax revenues in Iran. *Journal of Iran's Economic Essays (JIEE)*, 18(35), 169-191. [In Persian].
37. Salami, H., & Mafi, H. (2018). Predicting Export prices of the Iranian Pistachio Based on Commercial Cycles: Application of Structural Time Series Model. *Iranian Journal of Agricultural Economics and Development Research*, 49(4), 559-571. [In Persian].
38. Sevim, C., Oztekin, A., Bali, O., Gumus, S., & Guresen, E. (2014). Developing an early warning system to predict currency crises. *European Journal of Operational Research*, 237(3), 1095-1104.

39. Shiranifakhr, Z., Khoshakhlagh, R., & Sharifi, A. M. (2014). Estimating Demand Function for Natural Gas in the Industrial Sector of Iran using Structural Time Series Model (STSM). *Journal of Applied Economics Studies in Iran*, 3(11), 129-157. [In Persian].
40. Thayyib, P. V., Thorakkattle, M. N., Usmani, F., Yahya, A. T., & Farhan, N. H. (2023). Forecasting Indian Goods and Services Tax revenue using TBATS, ETS, Neural Networks, and hybrid time series models. *Cogent Economics & Finance*, 11(2), 2285649.
41. Xu, C., & Kong, Y. (2024). Random forest model in tax risk identification of real estate enterprise income tax. *Plos One*, 19(3), e0300928.
42. Zareie, P., Jalaee, S. A., & Sadeghi, Z. (2019) Simulation and prediction of the green tax effect on energy consumption and intensity in Iran using a genetic algorithm. *J Tax Res*, 27(42), 103-125. [In Persian].
43. Zhang, D., Zheng, S., & Fu, W. (2023). Research on the prediction model of Chinese tax revenue based on GM (1, 1) and LSSVM. *Information Technology and Control*, 52(4), 811-818.

استناد به این مقاله: دباغ، رحیم، عزتی، احمد و شعبانپور، راحله. (۱۴۰۴). مقایسه عملکرد الگوریتم‌های یادگیری ماشین مبتنی بر هوش مصنوعی با الگوی فضا حالت در پیش‌بینی درآمدهای مالیاتی استان آذربایجان غربی. پژوهشنامه مالیات، ۶۸(۳۳)، ۱۶۹-۱۲۵.



Journal of Tax Research is licensed under a Creative Commons Attribution-NonCommercial4.0 International License.