Revista Colombiana de Computación
Vol. 27, No. 1 January – June 2026, pp. 1-22
e-ISSN: 2539-2115, https://doi.org/10.29375/25392115.5326

Artificial intelligence for financial budget predictions: Neural networks lead the way

Inteligencia artificial para la predicción de presupuestos financieros: Las redes neuronales marcan el camino

Patricia Jimbo Santana[1,2,3,4], Erwis Melchor Pérez[5], Monica Jimbo Santana[1,3]

[1] Central University of Ecuador, Quito, Ecuador.
[2] University of Panama, Coclé Regional University Center, Postdoctoral Fellowship.
[3] Governance and Artificial Intelligence Research Group, Central University of Ecuador.
[4] Artificial Intelligence Research Group, Bolivarian University of Ecuador.
[5] Instituto Nacional de Astrofísica Óptica y Electrónica, Puebla, México.

prjimbo@uce.edu.ec, emelchor@inaoep.mx, djimbo@uce.edu.ec

(Received: 23 March 2025; accepted: 2 June 2026; published online: 30 June 2026)

Abstract. Predicting financial budgets remains an open challenge. Most studies focus on high-frequency markets, such as stock price prediction, leaving the budget domain aside. This work addresses that gap. We evaluate four artificial intelligence techniques for budget prediction: decision trees, random forests, linear regression, and multilayer perceptron. Additionally, we implement a hybrid version, MLP_GA, where the network hyperparameters are optimized using a genetic algorithm. For training and validation, we use two original datasets not publicly available: quarterly financial statements from Mexican entities, 2010-2024, and monthly financial statements from Ecuadorian organizations, 2019-2022. Both present characteristics typical of emerging economies, including episodes of high inflation, currency devaluations, fiscal policy changes, and the COVID-19 crisis. The results show that MLP_GA achieves superior predictive performance compared to the standard multilayer perceptron and the other techniques. Genetic algorithm optimization enables exploring multimodal and non-convex error surfaces, finding hyperparameter configurations –number of hidden neurons, learning rate, L2 regularization coefficient, maximum iterations, and early stopping tolerance– that significantly reduce overfitting, a critical problem when dealing with short time series of only forty to sixty quarterly observations. Statistical tests confirm that the differences are significant. The contribution is threefold. First, regarding the application domain, we present the first systematic and comparative evaluation of hybrid models combining genetic algorithms with multilayer perceptrons specifically applied to financial budget prediction, distinguishing ourselves from existing studies that focus almost exclusively on high-frequency stock market prediction. Second, regarding the geographical context, we validate our models in two Latin American economies that the literature has systematically ignored: Mexico and Ecuador, regions that represent less than one percent of the datasets used in the literature according to recent meta-analyses. Third, regarding the methodological justification, we demonstrate that genetic optimization is particularly effective for moderate-sized data, where more complex architectures like LSTMs or transformers suffer from overfitting due to their high parameter count. The combination of monthly and quarterly analysis improves model robustness, allowing accurate performance in both short-term scenarios and broader projections. These findings suggest that the MLP_GA approach can improve financial budgeting in data-constrained contexts. The paper concludes with a discussion of limitations and future research lines.

Keywords: Financial budget prediction, decision trees, neural networks, linear regression, random forest, predictive analysis.

Resumen. La predicción de presupuestos financieros sigue siendo un reto sin resolver. La mayoría de los estudios se centran en mercados de alta frecuencia, como la predicción de cotizaciones bursátiles, dejando de lado el ámbito presupuestario. Este trabajo aborda esa laguna. Evaluamos cuatro técnicas de inteligencia artificial para la predicción presupuestaria: árboles de decisión, bosques aleatorios, regresión lineal y perceptrón multicapa. Además, implementamos una versión híbrida, MLP_GA, en la que los hiperparámetros de la red se optimizan mediante un algoritmo genético. Para el entrenamiento y la validación, utilizamos dos conjuntos de datos originales que no están disponibles públicamente: estados financieros trimestrales de entidades mexicanas, 2010-2024, y estados financieros mensuales de organizaciones ecuatorianas, 2019-2022. Ambos presentan características típicas de las economías emergentes, incluyendo episodios de alta inflación, devaluaciones monetarias, cambios en la política fiscal y la crisis de la COVID-19. Los resultados muestran que MLP_GA alcanza un rendimiento predictivo superior en comparación con el perceptrón multicapa estándar y las demás técnicas. La optimización mediante algoritmos genéticos permite explorar superficies de error multimodales y no convexas, encontrando configuraciones de hiperparámetros –número de neuronas ocultas, tasa de aprendizaje, coeficiente de regularización L2, número máximo de iteraciones y tolerancia de detención temprana– que reducen significativamente el sobreajuste, un problema crítico cuando se trabaja con series temporales cortas de sólo cuarenta a sesenta observaciones trimestrales. Las pruebas estadísticas confirman que las diferencias son significativas. La contribución es triple. En primer lugar, en cuanto al ámbito de aplicación, presentamos la primera evaluación sistemática y comparativa de modelos híbridos que combinan algoritmos genéticos con perceptrones multicapa aplicados específicamente a la predicción de presupuestos financieros, lo que nos distingue de los estudios existentes que se centran casi exclusivamente en la predicción bursátil de alta frecuencia. En segundo lugar, en cuanto al contexto geográfico, validamos nuestros modelos en dos economías latinoamericanas que la literatura ha ignorado sistemáticamente: México y Ecuador, regiones que representan menos del uno por ciento de los conjuntos de datos utilizados en la literatura según metaanálisis recientes. En tercer lugar, en lo que respecta a la justificación metodológica, demostramos que la optimización genética resulta especialmente eficaz con conjuntos de datos de tamaño moderado, en los que arquitecturas más complejas, como las LSTM o los transformadores, adolecen de sobreajuste debido a su elevado número de parámetros. La combinación de análisis mensuales y trimestrales mejora la robustez del modelo, lo que permite un rendimiento preciso tanto en escenarios a corto plazo como en proyecciones más amplias. Estos hallazgos sugieren que el enfoque MLP_GA puede mejorar la elaboración de presupuestos financieros en contextos con limitaciones de datos. El artículo concluye con un debate sobre las limitaciones y las líneas de investigación futuras.

Palabras clave: Previsión presupuestaria, árboles de decisión, redes neuronales, regresión lineal, bosque aleatorio, análisis predictivo.

1 Introduction

Predicting financial budgets is important for both the public and private sectors. Strategic planning, resource allocation, and ultimately the organization's survival depend on these predictions. Classical forecasting methods, such as the ARIMA models developed by Box and Jenkins, have been the standard tool for a long time. They work well when data are stationary, when temporal patterns are clear, and when there are not too many variables involved (Wilson, 2016). These models have advantages: they are interpretable, have a solid theoretical foundation, and do not require much computational power. That is why they remain a benchmark. The problem is that today's financial environment is not that simple. There are nonlinear relationships, changing volatility, many variables that influence each other, noise everywhere, and large volumes of data. Under these conditions, traditional linear models are not sufficient (Hyndman & Athanasopoulos, 2021).

Then artificial intelligence comes into play. More specifically, machine learning. These techniques have shown that they can find patterns where linear models see nothing (Goodfellow et al., 2016). Decision trees, random forests, support vector machines, and neural networks have all been used. They have been applied with increasing success in finance: stock price prediction, credit risk assessment, fraud detection, portfolio optimization (Li & Law, 2024). If one reviews the literature carefully, it becomes clear that most studies focus on high-frequency markets. Stocks, stock indices, exchange rates. There, the signal is weak, noise dominates, and predictions are made for minutes, hours, or days ahead (Sezer et al., 2020). Organizational budgets, on the other hand, have received very little attention. This is striking, considering how important they are for organizations.

Financial budgets are different. Their frequency is lower: quarterly or monthly, not intraday. They exhibit seasonality and long cycles linked to fiscal policies, electoral cycles, or multi-year strategic plans. Although they can suffer economic shocks, they are more stable than stock prices. The signal-to-noise ratio is more favorable. The variables that matter are fundamentally economic: tax revenues, programmed expenditure, and investment targets. Nothing to do with market sentiment or short-term technical indicators. This difference is not minor. A model optimized to predict Bitcoin every five minutes is not necessarily good at predicting the quarterly budget of a Mexican company. The reverse is also true. That is why specific research is needed.

Some figures help to understand the magnitude of the problem. In 2020, a systematic review analyzed 150 articles on financial time series forecasting with deep learning published between 2005 and 2019. Some 94.7% used stock price or stock index data. Only 5.3% addressed other financial domains. Of that small group, barely two articles touched on financial statements or budgets (Sezer et al., 2020). Another study, a bibliometric one, was published in 2021. It reviewed 348 articles on AI and finance between 2011 and 2021. Stock price prediction accounted for 32% of the papers. Risk assessment 18%. Fraud detection 12%. Portfolio optimization 10%. And budget prediction? Less than 3% (Goodell et al., 2021). The conclusion is clear: there is a real gap in the application of advanced artificial intelligence techniques to the budget domain.

But there is a second gap. A geographical one. It has gone almost unnoticed. Most datasets come from developed markets. The United States, with its NYSE and NASDAQ exchanges, the United Kingdom with the LSE, Japan with the TSE, and global indices such as the S&P 500 and the FTSE 100. A recent meta-analysis on the geographical distribution of data in financial prediction studies using AI showed that 54% of the papers use data from the United States, 23% from Western Europe, 15% from China and East Asia, 8% from the rest of the world –India, Brazil, Turkey, South Africa–, and less than 1% focus on Latin America, excluding Brazil (Liu, 2025). This is problematic. Emerging economies do not behave like developed ones. Exchange rate volatility is higher, inflation is chronic in some cases, regulatory frameworks are more unstable, and financial markets are shallower. Generalizing results obtained in New York or London to Mexico or Ecuador is risky.

This research seeks to close both gaps at the same time. We use two original datasets, not publicly available. The first consists of quarterly financial budgets from various public and private entities in Mexico, from 2010 to 2024. The second consists of monthly financial statements from Ecuadorian organizations between 2019 and 2022. These data present challenges that do not appear in standard datasets from developed economies: episodes of high inflation, strong currency devaluations, government changes that modified fiscal policies, and the COVID-19 crisis. By including these two Latin American contexts, we can assess the robustness of models when macroeconomic conditions change drastically. And we provide original empirical evidence for a region that the literature has neglected.

Why genetic algorithms? It is not an arbitrary choice. When optimizing hyperparameters of a neural network, the function to be minimized –prediction error– is not a smooth surface with a single valley. Rather, it has a complicated shape, full of hills and valleys, with many points where the error is low surrounded by zones where the error is high. Technically, this is called a multimodal and non-convex error surface (Zhang et al., 2025). And that is exactly what appears when fitting neural networks to financial data. A small change in the learning rate, in the number of neurons, or in regularization can produce a huge and unpredictable jump in performance. Traditional methods like grid search or random search explore locally and get trapped in shallow valleys. Genetic algorithms, in contrast, maintain an entire population of solutions that simultaneously explore different regions of the space. Thus, they can find deeper valleys that would otherwise go unnoticed. Moreover, they naturally handle different types of hyperparameters: integers like the number of neurons, continuous reals like the learning rate, and categorical variables such as the activation function. Because they work with populations, they can be parallelized, drastically reducing computation time. Finally, they are robust to noisy or non-differentiable error functions, something very common in financial problems where small perturbations in the input data produce erratic changes in prediction error (Mandal Das et al., 2025).

The contribution of this work is threefold. First, regarding the application domain, we present the first systematic and comparative evaluation of hybrid models combining genetic algorithms with multilayer perceptrons specifically applied to financial budget prediction. This distinguishes us from existing studies, which focus almost exclusively on high-frequency stock market prediction. Second, regarding the geographical context, we validate our models in two Latin American economies that the literature has systematically ignored: Mexico and Ecuador. We generate original empirical evidence on the behavior of these techniques under macroeconomic conditions very different from those of developed markets. Third, regarding the methodological justification, we demonstrate that genetic algorithm optimization is particularly effective for moderate-sized datasets, such as those typical of quarterly budgets. In these cases, more complex techniques like LSTM networks or transformers suffer from overfitting or require data volumes that do not exist (Gür et al., 2025). The results we present later show that the hybrid approach consistently outperforms other techniques in terms of accuracy and predictive stability. The remainder of the paper is organized as follows: Section 2 reviews related work and situates our contribution; Section 3 presents the methodology; Section 4 reports the experiments and results; Section 5 provides the discussion; and Section 6 closes with conclusions and future work.

2 Related work

In the last two decades, artificial intelligence applied to finance has changed a lot. At first, simple feedforward neural networks with a single hidden layer were used. The multilayer perceptron was the king. It was mainly applied to predict corporate bankruptcies and to calculate credit ratings. And it worked better than classical statistical models like logistic regression or discriminant analysis (Khattak et al., 2023). A good example is a 2024 study where they applied machine learning, including neural networks, to bankruptcy prediction in a sample of European companies. The authors found that neural networks outperformed traditional statistical methods in classification accuracy (Shetty et al., 2022). Over time, more complex architectures arrived: LSTMs, GRUs, temporal convolutional networks, and, more recently, transformers. These enabled significant advances in capturing long-term temporal dependencies, especially in stock price prediction, where temporal relationships matter a lot (Vaswani et al., 2017).

Within the world of financial time series, recurrent networks and especially LSTMs became dominant. The reason is that they handle long-term dependencies well and mitigate the vanishing gradient problem, something that affects simpler recurrent networks (Hochreiter & Schmidhuber, 1997). The systematic review we mentioned earlier concludes something important: LSTMs outperform multilayer perceptrons in problems with complex temporal dependencies, such as intraday stock prices. However, that advantage does not always hold when the periodicity is lower, such as daily, weekly, or monthly data, and especially when data volumes are limited (Sezer et al., 2020). This last point is key to justifying our methodological choice. Quarterly financial budgets typically have between forty and sixty points per time series. Not many data points. In these cases, complex models like LSTMs or transformers tend to overfit because they have too many parameters to estimate from too few observations. A well-regularized multilayer perceptron with optimized hyperparameters generalizes better out of sample (Dietterich, 2000).

Hyperparameter optimization is a critical step. It has received a lot of attention in the literature, and for good reason. Traditional methods like grid search or random search are simple and interpretable, but inefficient when the search space has many dimensions. Moreover, they do not guarantee finding optimal configurations in complex problems (Lin et al., 2025). A foundational 2012 paper in the Journal of Machine Learning Research showed that random search can be more efficient than grid search in high-dimensional spaces. But it still has limitations when the objective function has multiple local optima (Bergstra & Bengio, 2012). Bayesian optimization is more advanced and performs well in many problems, but it can also get trapped in local optima due to the multimodal nature of the loss function (Snoek et al., 2012). A comprehensive review published in 2023 in Swarm and Evolutionary Computation examined metaheuristics for hyperparameter optimization in neural networks. The conclusion was that evolutionary algorithms, and particularly genetic algorithms, offer clear advantages in global exploration of the search space and robustness against multimodality (Zhang et al., 2025).

Genetic algorithms have been applied with increasing success to neural network optimization in finance. An early but influential study, published in 2018 in Sustainability, optimized complete LSTM architectures using genetic algorithms. They optimized the number of hidden layers, neurons per layer, learning rate, and window size. They compared their approach with Bayesian optimization on fifteen different financial series. The genetic algorithm won in twelve out of fifteen (Chung & Shin, 2018). More recently, a 2025 paper in the proceedings of the CECCT 2024 conference integrated genetic algorithms with multilayer perceptrons for stock price prediction. They incorporated backtesting constraints into the fitness function (Sun et al., 2025). The improvements in signal-to-noise ratio reached 28.56% on average over models without optimization. Another 2025 article, this one in Heliyon, proposed a hybrid architecture combining genetic algorithms, attention mechanisms, and fuzzy logic. They called it GA-Attention-Fuzzy-Stock-Net. They applied it to technology stocks and reported improvements in the mean absolute percentage error of up to 34% compared to standard LSTM (Gülmez, 2025). That said, it is worth noting that so much architectural complexity, three optimization mechanisms at once, raises the question of whether it is necessary for problems with limited data, such as budget prediction. Sometimes less is more.

Despite all these advances, if one reviews the literature critically, one finds an overwhelming concentration on the stock market domain. That is the first gap. Of the 150 articles reviewed by Sezer and colleagues in 2020, 142 –94.7%– used stock price or stock index data. Only 8 articles addressed other financial domains (Sezer et al., 2020). This asymmetry is not a minor detail because stock market prediction and budget prediction are very different. In the stock market, frequency is high: minutes, hours, days. The signal-to-noise ratio is extremely low, often below 0.1. Unpredictable exogenous variables, such as news or market sentiment, have a dominant influence. There are multiple volatility regimes with volatility clustering. In contrast, budget prediction has low frequency: monthly, quarterly, and yearly. The signal-to-noise ratio is moderate, with clearly identifiable seasonal and trend components. The variables that matter are fundamentally predictable: GDP, inflation, and fiscal policies. And growth patterns are more stable, less prone to abrupt regime changes.

A recent paper that deserves attention because of its proximity to our research was published in 2025 in Panoeconomicus. The authors applied advanced AI models, including temporal fusion transformers, to public expenditure prediction in Turkey using quarterly data (Gür et al., 2025). They found that transformers capture complex interactions between variables but require massive data, thousands of time points. For typical budget series, which rarely exceed one hundred quarterly observations, they are impractical. That is why they explicitly recommend using more parsimonious models, such as a well-optimized multilayer perceptron. This reinforces our methodological justification. Another 2025 paper in PLOS ONE proposed an architecture called LTRNet that combines LSTMs, transformers, and residual networks for financial risk assessment. But they used datasets with more than one-hundred-thousand-time points from Kaggle and Yahoo Finance (Liu, 2025). This demonstrates that complex architectures are only applicable when massive data volumes are available, something that does not exist in the budget domain.

The second gap is geographical. It has also been systematically ignored. The 2025 meta-analysis quantified the geographical distribution of datasets in financial prediction studies using AI. The numbers are telling: 54% of the papers use data from the United States, 23% from Western Europe, 15% from China and East Asia, 8% from the rest of the world –India, Brazil, Turkey, South Africa–, and less than 1% focus on Latin America excluding Brazil (Liu, 2025). This skewed distribution raises serious questions about the external validity of results when applied to emerging and developing economies. These economies have higher exchange rate volatility, chronic inflation in some cases, more unstable regulatory frameworks, and shallower financial markets. To date, there is no systematic study evaluating hybrid models combining genetic algorithms with multilayer perceptrons for budget prediction in Latin American economies. This is the contribution of this research.

3 Methodology

The methodology we propose seeks to systematically and reproducibly evaluate how different artificial intelligence techniques behave when applied to financial budget prediction. We make an important distinction: on one hand, we train models with the default hyperparameters that come with the libraries, and on the other hand, we optimize those hyperparameters using genetic algorithms. Before going into details, we need to clarify something about accounting accounts. The chart of accounts is not the same across countries, and financial statements present aggregate accounts at different levels. Not all accounts end up in the predictive models. Why? First, because not all of them significantly impact the overall financial situation of the institution. We are interested in those items that concentrate most of the variability and are relevant for decision-making. Second, predicting all accounts would require an enormous amount of data and computational resources. The additional cost and effort are not always justified, especially when working with quarterly data where the number of observations is limited. Third, some accounts have insufficient or inconsistent historical data. Outliers, missing values, or changes in accounting recording criteria over time may appear. And fourth, financial models need a balance between accuracy and simplicity. Too many variables make the model excessively complex, difficult to interpret, and more prone to overfitting.

For all these reasons, the analysis focuses on critical accounts for evaluating financial performance. We work mainly with total assets, total liabilities, and net income –the institution's profit or loss over a period–. Why these three? Total assets reflect the size of the entity and its capacity to generate resources. Total liabilities show the level of debt and the financing structure. Net income synthesizes operational efficiency and profitability. Together, these three variables provide a comprehensive view of financial health according to generally accepted accounting principles. Predicting them allows managers to anticipate liquidity, solvency, or profitability problems before they materialize.

The study has two parts. In the first, we evaluate each AI technique individually using the default hyperparameters that come with standard Python libraries. In the second, we take those same techniques –with special emphasis on the multilayer perceptron– and combine them with genetic algorithms to optimize their hyperparameters during training. We use four AI techniques for regression problems: decision tree, random forest, linear regression, and multilayer perceptron. The latter is further optimized with a genetic algorithm that searches for the best combination of hyperparameters. We call the resulting model MLP_GA. All implementations were done with standard Python libraries: scikit-learn for the base models and DEAP for the genetic algorithm (De Rainville et al., 2012).

Now, let us talk about the multilayer perceptron architecture in detail, because without this nobody could reproduce our results. The MLP we implemented has an input layer with as many neurons as predictor variables included in the model; we detail this in the data section. Then comes a single hidden layer. The number of neurons in that layer is precisely one of the hyperparameters we optimize with the genetic algorithm. Why only one hidden layer and not more? Because previous research on low-frequency financial time series shows that deeper architectures tend to overfit when there are few observations. And that is exactly our case: quarterly budgets give us between forty and sixty points per series. In the hidden layer, we use the hyperbolic tangent as the activation function. It gives outputs between -1 and 1 and has shown good performance in financial regression problems. The output layer has a single neuron –the predicted value– and uses linear activation, which is most suitable for regression because the dependent variable can take any real value. Training is done with error backpropagation, adjusting the weights to minimize the mean squared error between what the network predicts and the actual value. Training stops when we reach the maximum number of iterations or when the improvement in validation error falls below a threshold. Both parameters are also optimized.

Let us move on to the genetic algorithm. For other researchers to reproduce what we did, we need to specify its configuration precisely. We follow Holland's canonical scheme, with the usual extensions. The initial population has fifty individuals. Each individual represents a complete combination of MLP hyperparameters, coded as a vector mixing real and integer values. An important point, worth clarifying because it could be misinterpreted: the genetic algorithm does NOT evaluate all possible combinations of hyperparameters. That would be unfeasible. The hyperparameter space is continuous in some cases and combinatorially enormous in others. An exhaustive exploration is impossible. Instead, the algorithm maintains a population of candidate solutions that evolves generation after generation through selection, crossover, and mutation. This way, it explores promising regions without needing to evaluate everything.

What hyperparameters do we optimize? First, the number of neurons in the hidden layer is an integer value between five and one hundred. Second, the backpropagation learning rate is a continuous real number between 0.0001 and 0.1. Third, the L2 regularization coefficient –also called weight decay–, which helps prevent overfitting, has real values between 0.00001 and 0.01. Fourth, the maximum number of training iterations is between one hundred and one thousand. Fifth, the tolerance for early stopping, which defines the minimum improvement in validation error required to continue training. The objective function that the genetic algorithm minimizes is the mean squared error obtained from five-fold cross-validation. The chosen configuration is the one that gives the lowest estimated generalization error.

The operational parameters of the genetic algorithm are as follows. The crossover probability between two parents is 0.9: 90% of parent pairs produce offspring. We use uniform crossover, where each hyperparameter of the child is inherited from one of the two parents with equal probability. The mutation probability per gene is 0.1: each hyperparameter has a 10% chance of being randomly modified. For continuous variables, mutation adds Gaussian noise with zero mean and standard deviation proportional to the hyperparameter's range. For integer variables, we apply a random shift within the allowed limits. Selection is by tournament, with tournament size three: three individuals are chosen at random, and the one with the best objective function value is selected as a parent. The algorithm evolves for fifty generations. By then, the population has usually converged to a promising region. At the end, we select the individual with the best objective function value across the entire evolutionary history, that is, the hyperparameter combination that produced the lowest cross-validation error.

A crucial methodological point. In financial prediction, especially with quarterly or monthly data that have an explicit temporal structure, one cannot perform a random split of the observations. That is a serious mistake. Random splitting works when observations are independent, but in time series, it generates data leakage: it allows future observations to be used to predict past observations during training. Performance estimates become artificially optimistic and do not hold up in practice. To avoid this, we use sequential temporal splitting, also known as out-of-time validation. We order the data chronologically, from oldest to newest. We assign the oldest 80% of observations to the training set and the most recent 20% to the test set. This way, the model is trained only with past information to predict the future, which is exactly what happens when the model is used in real life. For the Mexican quarterly data, 2010 to 2024, we train with quarters from the first quarter of 2010 to the fourth quarter of 2021, and test with quarters from the first quarter of 2022 to the fourth quarter of 2024. For the Ecuadorian monthly data, January 2019 to December 2022, we train with months from January 2019 to August 2021, and test with months from September 2021 to December 2022.

Data preprocessing involves several steps. First, we visually and statistically inspect each time series to identify outliers. We define an outlier as any point that is more than three standard deviations from the series moving average. Detected outliers are imputed by linear interpolation between the nearest temporal neighbors. This preserves the temporal structure of the series. Second, we verify stationarity using the augmented Dickey-Fuller test. If the series is non-stationary, we apply first-order differencing to remove the trend. Third, we decompose each series into trend, seasonality, and residual components using the STL seasonal decomposition method. This allows us to separately analyze the different temporal patterns. Fourth, we apply a logarithmic transformation to series that exhibit increasing variance with the level, a very common phenomenon in finance where fluctuations grow proportionally to the series value. Fifth, we apply min-max normalization to scale all series to the range between zero and one, using the formula $X_{norm} = (X - X_{min}) / (X_{max} - X_{min})$. This normalization is necessary because AI techniques, particularly the multilayer perceptron, are sensitive to the scale of input variables. If one variable has a much larger range than another, it can dominate the learning process.

We use cross-validation during training to estimate the performance of different hyperparameter configurations without touching the final test set. We employ five-fold cross-validation, but with an important modification to respect the temporal structure, with no random folds. We use rolling window cross-validation, where each training fold contains only observations that precede those in the validation fold. Specifically, for a training set containing T temporally ordered observations, we create five consecutive partitions. The first fold uses the first $T/5$ observations for training and the next $T/5$ for validation. The second fold uses the first $2T/5$ observations for training and the next $T/5$ for validation. And so on. This strategy is called expanding window cross-validation. It respects the temporal order and provides a more realistic estimate of generalization error than standard cross-validation with random folds.

Figure 1 shows the methodology for evaluating AI techniques without hyperparameter optimization. The first step is preprocessing: outlier detection and imputation, stationarity verification with differencing if needed, seasonal decomposition, logarithmic transformation when applicable, and min-max normalization. The second step is temporal data splitting, strictly respecting the chronological order, where historical data are used for training and the most recent periods for evaluation, avoiding data leakage. The third step is training. We take the default hyperparameters from scikit-learn and train the model using five-fold temporal cross-validation with a rolling window. The result is a model trained with default hyperparameters. The fourth step is evaluation. We apply the trained model to the test set to generate predictions, and we measure quality using the metrics described below, reporting both point estimates and confidence intervals obtained via bootstrap.

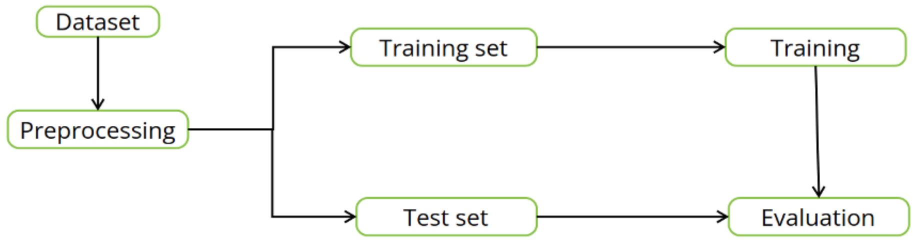

Figure 1. Methodology used without hyperparameter optimization

Figure 2 describes the methodology for evaluating AI techniques with hyperparameter optimization using genetic algorithms. The variation from the previous methodology is substantial and concentrates on the third step, training. In this case, before training the final model, we run the genetic algorithm with all the configurations detailed above: an initial population of fifty individuals, each representing a combination of MLP hyperparameters, evaluated using five-fold temporal cross-validation, with uniform crossover at probability 0.9, Gaussian mutation at probability 0.1 per gene, tournament selection of size three, and evolution for fifty generations. At the end of the evolutionary process, we select the hyperparameter combination that produced the lowest mean squared error averaged across the cross-validation folds. With these optimal hyperparameters, we retrain the multilayer perceptron, this time using the entire training set, and obtain the final model. Again, it is important to clarify that the genetic algorithm does not evaluate all possible hyperparameter combinations. It explores the search space adaptively, evaluating only the combinations corresponding to the individuals generated over the generations. The result of this step is a model trained with hyperparameters optimized by the genetic algorithm. In the fourth step, we evaluate the model with optimized hyperparameters on the test set using the same quality metrics as in the previous methodology, allowing a fair comparison between the optimized model and the baseline models.

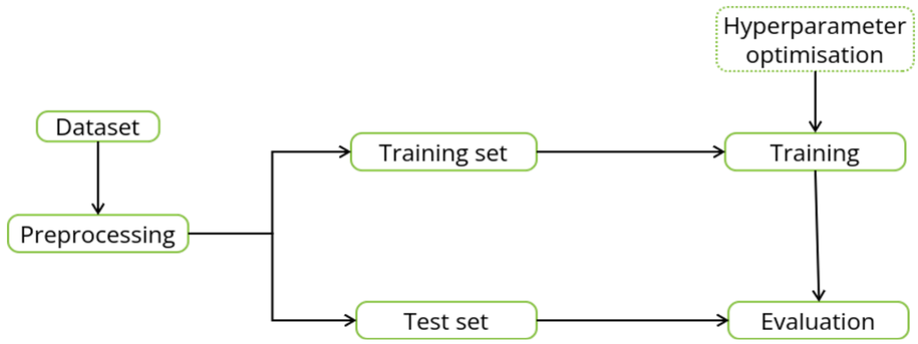

Figure 2. Methodology used with hyperparameter optimization

4 Experiments and results

For our experimental study on the evaluation and prediction of financial budgets, we selected the following balance sheet accounts: assets, liabilities, and additionally, net income or loss from the income statement.

For the implementation, four financial institutions were considered, each with authorized operating levels I, II, III, and IV. In the case of Mexican institutions, they are authorized by the National Banking and Securities Commission – CNBV to carry out operations. Quarterly data from the period between 2010 and 2024 were analyzed. The information analyzed is available on the CNBV website on a quarterly basis. Additionally, monthly data from an Ecuadorian financial institution with information from 2019 to 2022 were considered, totaling 608 accounts, authorized by the Superintendency of Banks – SB.

To evaluate the quality of the AI techniques applied to the time series problem, the following quality metrics commonly used in the literature were employed: Mean Squared Error – MSE Equation (1), Coefficient of Determination – R^2 Equation (2) and Mean Absolute Error – MAE Equation (3).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \bar{y}| \quad (3)$$

MSE penalizes large errors, while MAE provides a more robust measure against outliers. The R^2 coefficient, for its part, allows evaluating the model's ability to explain the variability of the time series. The combined use of these metrics enables a more complete assessment of predictive performance.

Table 1 presents the accounting accounts classified by operational levels, where these levels correspond to the degree of operation of the financial institution, ordered from lowest to highest. Thus, Level I represents the most basic operational level, while Level IV reflects the highest level. This classification allows analyzing the structural evolution of the main financial accounts according to operational level, providing a framework for interpreting their aggregate behavior. The income statement reflects the revenues, expenses, and net income –profit or loss– of an institution over a specific period. In this context, its prediction requires considering those accounting accounts that representatively capture the operational and financial activity of the institution. As observed in Table 1, there is an increasing trend in the values of assets, liabilities, and the income statement as the operational level of the institution increases. This pattern suggests a structural relationship between the operational level and the magnitude of the analyzed financial variables. This relationship is relevant for the forecasting problem, as it allows inferring how financial variables might evolve in response to changes in operational level, providing useful information for strategic planning and decision-making.

Table 1. Aggregate values of assets, liabilities, and the income statement by operational level of financial institutions

	Operational Level			
Account	Level I	Level II	Level III	Level IV
Asset	426,533	171,812	2,293,168	5,781,968
Liability	333,705	129,298	1,879,322	4,178,814
Income Statement	6,925	2,948	43,909	49,841

Five financial institutions representing different operational levels were analyzed, allowing evaluation of model behavior under different scales and financial complexities. Table 2 shows the number of records analyzed, as well as the number of accounting accounts each institution possesses.

Table 2. Distribution of the number of accounting accounts and records by financial institution and operational level

Database	# of accounts	# of first level accounts	# of second level accounts	# of third level accounts	# of fourth level accounts	Institution level
Institution I	**610**	**177**	**214**	**140**	**49**	**-**
Institution II	67	19	21	27	0	I
Institution III	67	19	21	27	0	II
Institution IV	67	19	21	27	0	III
Institution V	67	19	21	27	0	IV

Figure 3 and Figure 4 compare the results obtained by each artificial intelligence technique in predicting the Asset and Liability accounts, respectively. In both cases, the models were evaluated using default hyperparameter configurations, allowing a baseline comparison between the different techniques. The comparison was performed using the actual values from the test set, corresponding to the final evaluation period of the financial series. It is observed that the MLP model optimized with genetic algorithms –MLP_GA– presents a better fit to the actual values compared to the other models, evidenced by a smaller deviation between the predicted and observed series. The brown line represents the actual values of the institution's accounts, while the other lines correspond to the predictions generated by each AI technique. The monthly-level analysis allows capturing short-term variations in financial series, which is relevant for evaluating the models' ability to adapt to dynamic changes in the institution's economic behavior.

Furthermore, Figure 5 and Figure 6 present the results at the quarterly level, in accordance with the availability of information published by financial institutions on the CNBV portal. This level of aggregation allows analyzing the behavior of financial series from a more stable perspective, reducing the short-term variability observed in the monthly analysis and facilitating the identification of medium-term trends. In this context, it is observed that the models maintain performance patterns consistent with those obtained in the monthly analysis, with the MLP_GA model again standing out for its greater ability to approximate the trend of actual values. The quarterly analysis is particularly relevant for evaluating the models' generalization ability, as it allows verifying their performance over wider time horizons that are less sensitive to point fluctuations.

The results obtained show that Decision Tree – DT and Random Forest – RF models present lower performance in predicting the financial institution's balance sheet compared to the other approaches evaluated. These techniques were implemented using default hyperparameters and attempts to improve them through Genetic Algorithm – GA optimization did not generate significant increases in their performance. On the other hand, the Multilayer Perceptron – MLP model with default configuration, represented by the orange line, shows a better approximation to the actual values compared to DT and RF, evidencing its greater flexibility for modeling nonlinear relationships in financial series. Based on this result, we proceeded to optimize the MLP hyperparameters using a genetic algorithm, which improved its predictive capacity, reducing the deviation from observed values and capturing the series trend more accurately. Consequently, the optimized MLP model –MLP_GA– presents the best performance among the evaluated techniques, both in terms of visual fit and the quantitative metrics analyzed. However, this superiority should be interpreted within the context of the data used and is supported by the comparative analyses presented in the corresponding sections.

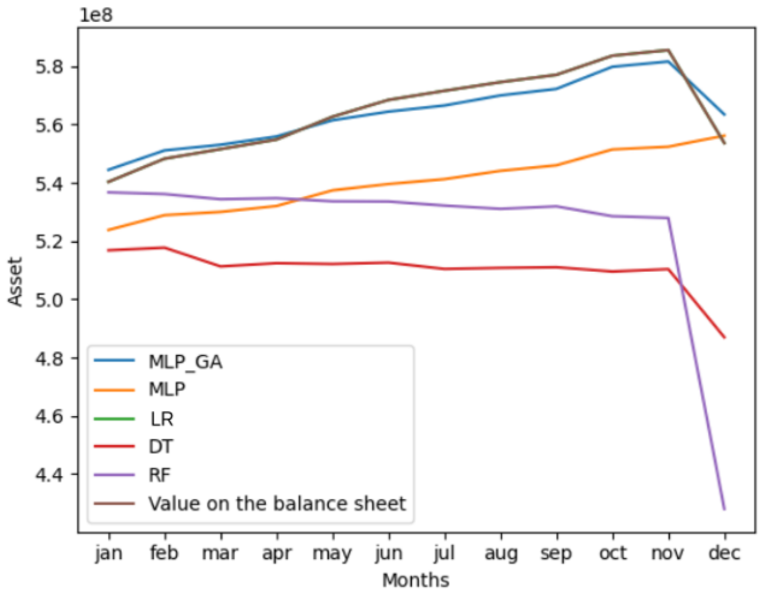

Figure 3. Comparison between the actual values from the test set and the predicted values for the Assets account obtained using artificial intelligence models with default hyperparameter settings for Institution I

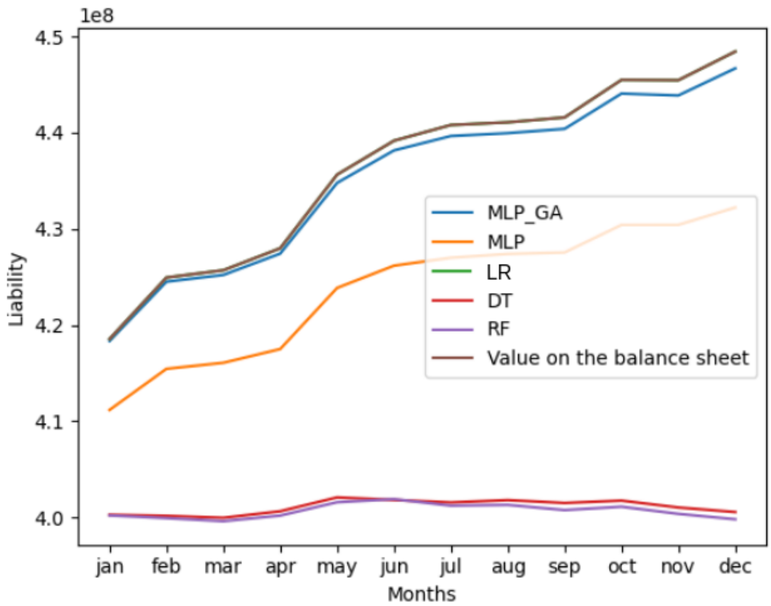

Figure 4. Comparison between the actual values from the test set and the predicted values for the Liabilities account obtained using artificial intelligence models with default hyperparameter settings for Institution I

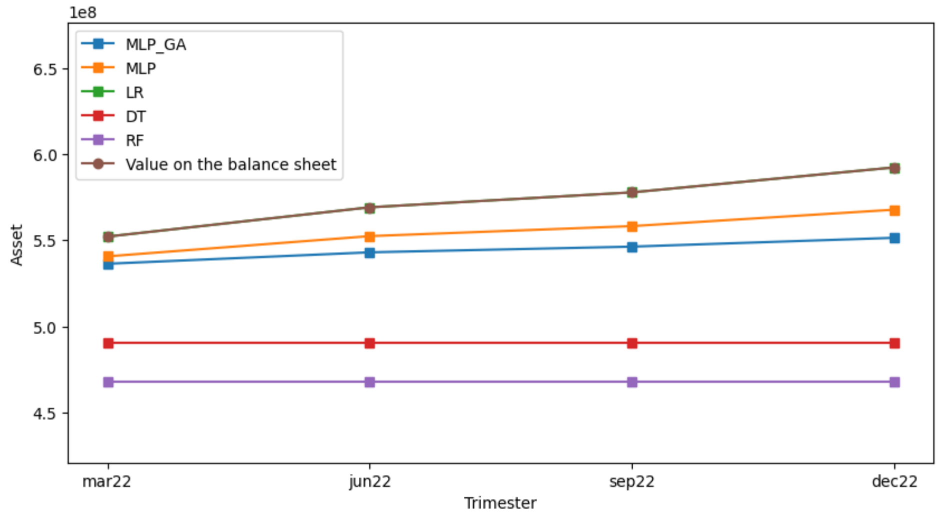

Figure 5. Comparison of actual values from the test set and predicted values for the Assets account using artificial intelligence models with default hyperparameters –quarterly analysis– for Institution I.

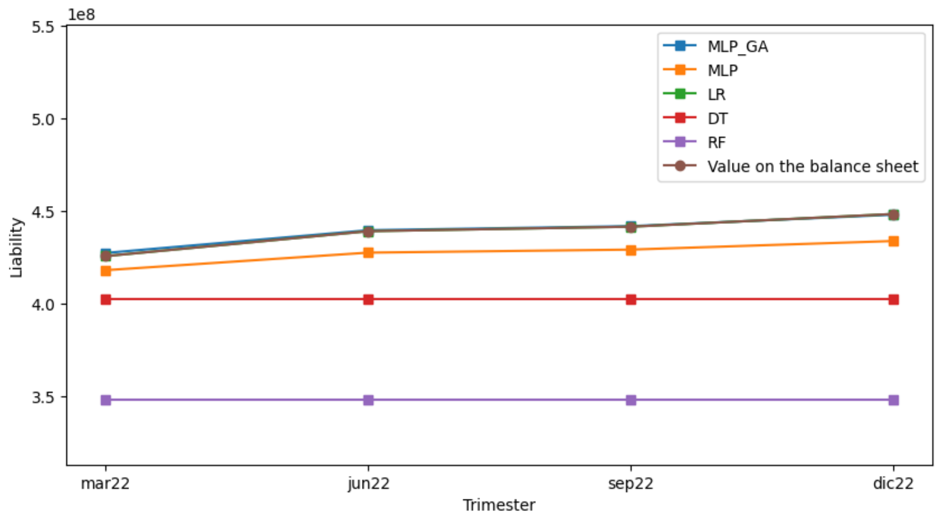

Figure 6. Comparison of actual values from the test set and predicted values for the Liabilities account using artificial intelligence models with default hyperparameters –quarterly analysis– for Institution I.

5 Discussion

Figure 7 shows the evolution of the profit or loss account from the income statement, which represents the difference between revenues and expenses of the financial institution. The results indicate that the artificial intelligence models manage to approximate the trend of the time series, with the MLP model optimized by genetic algorithms –MLP_GA– presenting the smallest deviation from the actual values. The monthly analysis allows evaluating the models' ability to capture short-term variations in financial performance. Figure 8 presents the results at the quarterly level. In this case, a more stable behavior is observed, which allows for analyzing the models' ability to capture medium-term trends. Together, both analyses show that the models are capable of adequately representing the temporal dynamics of the variable, with MLP_GA offering the best performance under the evaluated conditions.

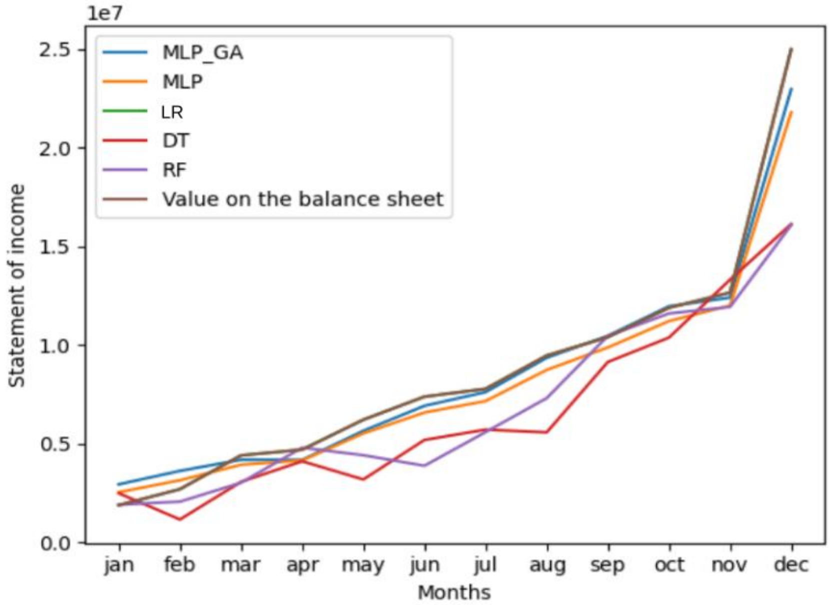

Figure 7. Comparison of actual and predicted values of the profit or loss account from the income statement using artificial intelligence models for Institution I

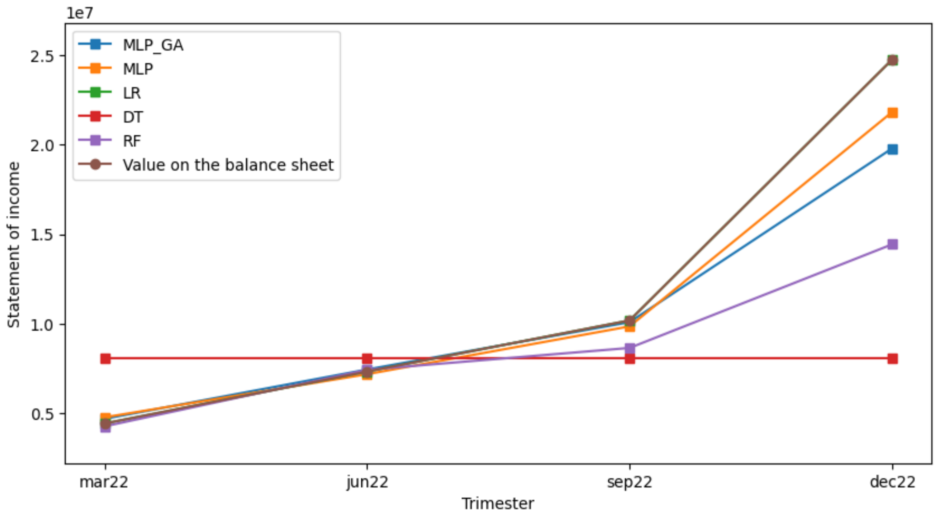

Figure 8. Comparison of actual values from the test set and predicted values of the profit or loss account from the income statement using artificial intelligence models –quarterly analysis– for Institution I

Figure 9 shows the mean squared error obtained with the AI techniques. The results show that the MLP model optimized by genetic algorithms –MLP_GA– presents the lowest error values, indicating greater prediction accuracy compared to the other models evaluated. It is observed that MLP_GA predictions show less dispersion around the actual values, while the remaining models show higher errors. Likewise, this behavior is consistent when considering the different evaluation metrics used in the study –MSE, MAE, and R^2–, which reinforces the robustness of the MLP_GA model in predicting financial statements under the analyzed conditions.

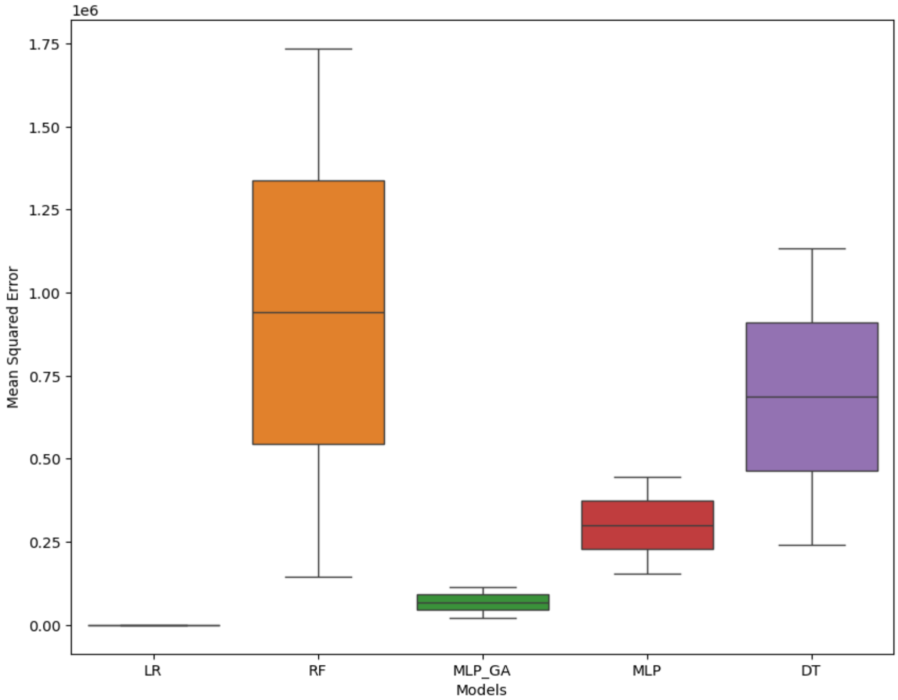

Figure 9. Comparison of Mean Squared Error – MSE among artificial intelligence models for financial statement prediction for Institution I.

Table 3 presents the comparative values obtained by each AI technique against the actual value of the account on the balance sheet, expressed in US dollars. As observed, the Linear Regression – LR model produces values that are very similar to each other and close to an average trend, which suggests a limited capacity to capture the temporal dynamics of the analyzed financial series. This behavior can be explained by the linear nature of the model, which restricts its ability to model the non-linear relationships present in the data. In contrast, the MLP model optimized by genetic algorithms –MLP_GA– presents values closer to those observed, evidencing a better fitting capacity. This model was configured using the tanh activation function and optimized using a sphere-type objective function Equation (4), with population parameters and number of generations set to 12, in accordance with the annual periodicity of the data. These results are consistent with the evaluation metrics presented previously, which support the better performance of the MLP_GA model under the analyzed conditions.

$$f(x) = \sum_{i=1}^{n} x_i^2 \tag{4}$$

Figure 10 shows the prediction of the Assets account for Institution II at Level I. The results show that the evaluated models manage to approximate the trend of the time series, although with different levels of precision. In particular, the MLP model optimized by genetic algorithms –MLP_GA– presents the smallest deviation from the actual values, evidencing a better fit compared to the other techniques. This analysis is carried out at the quarterly level, which allows evaluating the models' performance in capturing medium-term trends, reducing the influence of short-term fluctuations present in financial data.

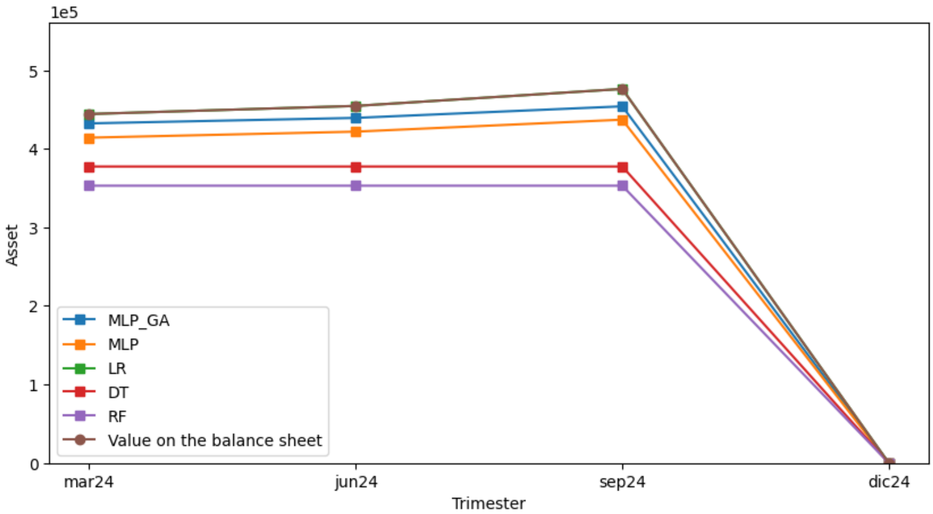

Figure 10. Comparison of actual and predicted values of the Assets account for Institution II at Level I using artificial intelligence models

Table 3. Comparison of actual and predicted balance sheet values using artificial intelligence models –values in USD– for Institution I.

LR [$]	RF [$]	MLP_GA [$]	MLP [$]	DT [$]	Balance [$]	Period
28,174,914.50	27,985,744.35	28,389,677.80	27,313,131.86	26,948,753.54	28,174,914.50	January
28,588,577.32	27,954,783.13	28,736,143.68	27,578,019.72	26,995,807.70	28,588,577.32	February
28,759,191.32	27,862,307.71	28,837,827.34	27,635,722.09	26,660,462.78	28,759,191.32	March
28,930,726.08	27,880,576.95	28,985,583.09	27,742,152.50	26,718,142.11	28,930,726.08	April
29,339,536.83	27,825,115.72	29,275,797.75	28,021,384.02	26,704,103.15	29,339,536.83	May
29,639,873.51	27,821,472.89	29,430,557.66	28,135,078.54	26,728,156.75	29,639,873.51	June
29,803,139.07	27,749,815.81	29,539,285.20	28,222,235.58	26,615,047.29	29,803,139.07	July
29,957,956.96	27,691,041.29	29,720,315.97	28,370,398.68	26,634,258.02	29,957,956.96	August
30,088,490.77	27,735,741.85	29,837,290.77	28,469,143.54	26,646,172.37	30,088,490.77	September
30,431,223.75	27,558,768.82	30,233,022.95	28,753,693.30	26,569,316.49	30,431,223.75	October
30,531,375.71	27,527,412.35	30,327,198.30	28,802,845.80	26,612,076.38	30,531,375.71	November
28,868,582.98	22,321,146.78	29,379,594.43	28,999,989.11	25,392,247.12	28,868,582.98	December

Figure 11 shows the results, which indicate that the evaluated models manage to capture the general trend of the time series, although with differences in their fitting capacity. In particular, the MLP model optimized by genetic algorithms –MLP_GA– shows the smallest deviation from the actual values, evidencing better predictive performance compared to the other techniques. The analysis is carried out at the quarterly level, which allows evaluating the models' ability to represent medium-term trends, reducing the effect of short-term variations in financial series.

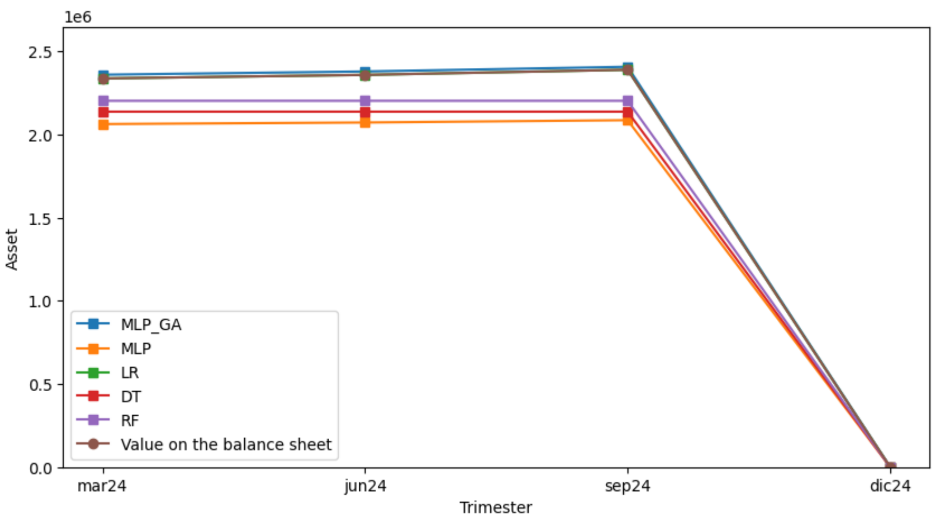

Figure 11. Comparison of actual and predicted values of the Assets account for Institution III at Level II using artificial intelligence models

Figure 12 shows that although all models manage to follow the general trend of the time series, there are differences in prediction accuracy. In particular, the MLP model optimized by genetic algorithms –MLP_GA– presents the best fit, showing a smaller deviation from the actual values. The analysis is carried out at the quarterly level, which allows evaluating the models' ability to capture more stable patterns over time and analyzing their performance in modeling medium-term trends.

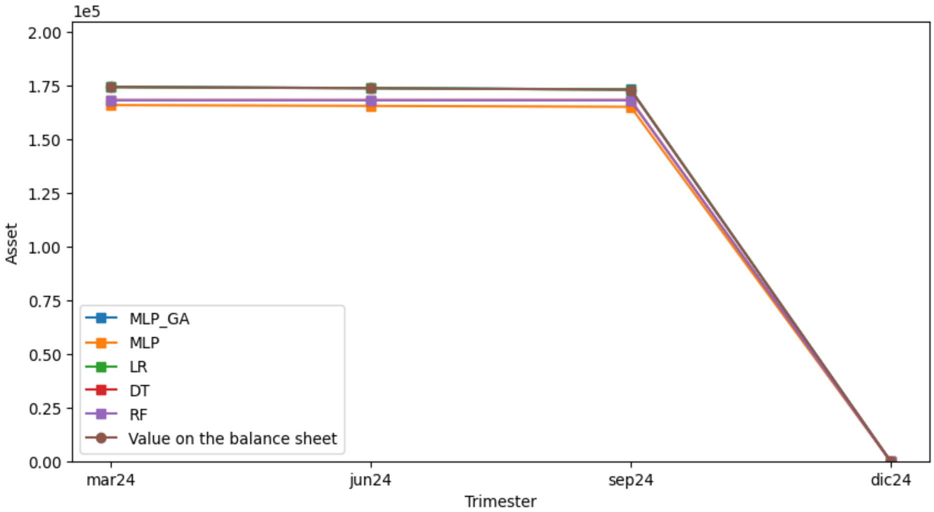

Figure 12. Comparison of actual and predicted values of the Assets account for Institution IV at Level III using artificial intelligence models

Figure 13 shows the results, which indicate that the evaluated models can reproduce the general trend of the time series, although with differences in their level of accuracy. In particular, the MLP model optimized by genetic algorithms –MLP_GA– shows the best performance, presenting the smallest deviation from the actual values. The analysis is carried out at the quarterly level, which allows evaluating the models' ability to capture more stable and consistent trends over time, characteristic of medium-term horizons.

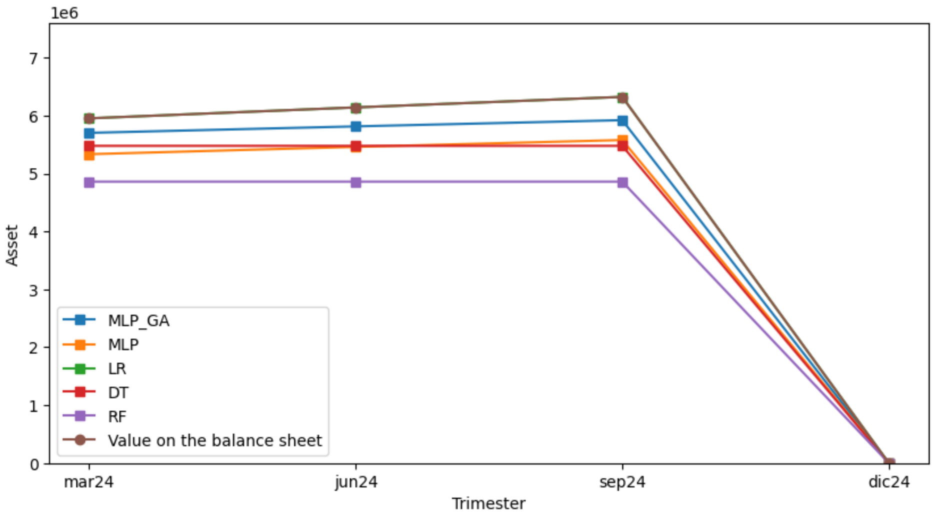

Figure 13. Comparison of actual and predicted values of the Assets account for Institution V at Level IV using artificial intelligence models

Table 4 shows the comparison of actual and predicted balance sheet values using artificial intelligence models for a Level I financial institution, for quarterly periods corresponding to the year 2022. As observed, there are differences in prediction accuracy among the evaluated models. In particular, the MLP model optimized by genetic algorithms –MLP_GA– presents values closer to those observed, indicating better predictive performance compared to the other techniques. These results are consistent with the graphical analysis and the quantitative metrics previously presented, which reinforce the model's ability to capture the dynamics of the time series under real conditions.

Table 4. Comparison of actual and predicted balance sheet values using artificial intelligence models for a Level I financial institution, for quarterly periods corresponding to the year 2022

LR [$]	RF [$]	MLP_GA [$]	MLP [$]	DT [$]	Balance [$]	Period
552,088,089.12	467,699,927.32	536,326,403.32	540,608,768.50	490,344,852.29	552,088,089.12	mar22
569,092,729.37	467,699,927.32	542,959,035.75	552,360,623.21	490,344,852.29	569,092,729.37	jun22
577,768,250.94	467,699,927.32	546,267,871.96	558,181,691.66	490,344,852.29	577,768,250.94	sep22
592,254,906.65	467,699,927.32	551,394,051.58	567,714,952.22	490,344,852.29	592,254,906.65	dec22

Table 5 presents the comparison of actual and predicted balance sheet values using artificial intelligence models for a Level I financial institution, for quarterly periods corresponding to the year 2024. As with the previous results, differences in prediction accuracy are observed among the evaluated models. In particular, the MLP model optimized by genetic algorithms –MLP_GA– presents values closer to those observed, indicating better predictive performance compared to the other techniques. This consistent behavior across different periods reinforces the model's stability and its ability to generalize to data not used during training. Likewise, these results are consistent with the error metrics previously analyzed, which supports the robustness of the proposed approach.

Table 5. Comparison of actual and predicted balance sheet values for a Level I financial institution in 2024

LR [$]	RF [$]	MLP_GA [$]	MLP [$]	DT [$]	Balance [$]	Period
444,367.00	353,103.85	432,423.00	414,246.82	377,533.80	444,367.00	mar24
454,651.00	353,103.85	439,414.49	421,906.90	377,533.80	454,651.00	jun24
476,286.00	353,103.85	454,142.57	437,177.51	377,533.80	476,286.00	sep24

Table 6 presents the comparison of actual and predicted balance sheet values using artificial intelligence models for a Level II financial institution, for quarterly periods in 2024. In line with the results obtained in previous operational levels, differences in prediction accuracy are observed among the evaluated models. In particular, the MLP model optimized by genetic algorithms –MLP_GA– presents values closer to those observed, indicating better predictive performance compared to the other techniques. This consistent behavior across different operational levels suggests that the model maintains its generalization capability in scenarios with different financial structures. Likewise, the results are consistent with the evaluation metrics previously analyzed, which reinforces the robustness of the proposed approach.

Table 6. Comparison of actual and predicted balance sheet values for a Level II financial institution in 2024

LR [$]	RF [$]	MLP_GA [$]	MLP [$]	DT [$]	Balance [$]	Period
174,297.00	168,116.30	174,363.20	165,865.45	168,215.00	174,297.00	mar24
173,756.00	168,116.30	173,817.84	165,512.85	168,215.00	173,756.00	jun24
173,096.00	168,116.30	173,152.41	165,082.69	168,215.00	173,096.00	sep24

Table 7 presents the comparison of actual and predicted balance sheet values using artificial intelligence models for a Level III financial institution, for quarterly periods in 2024. Similarly, it can be observed that the MLP_GA technique provides the best prediction. The results show consistent behavior with the operational levels previously analyzed, where differences in prediction accuracy are observed among the models. In particular, the MLP model optimized by genetic algorithms –MLP_GA– presents the closest values to those observed, reflecting better performance in estimating the financial series. This consistency across different operational levels reinforces the model's ability to adapt to more complex financial structures. Likewise, the results align with the evaluation metrics previously reported, which supports the robustness of the proposed approach.

Table 7. Comparison of actual and predicted balance sheet values for a Level III financial institution in 2024

LR [$]	RF [$]	MLP_GA [$]	MLP [$]	DT [$]	Balance [$]	Period
2,338,470.00	2,203,903.15	2,360,889.31	2,063,889.69	2,138,411.00	2,338,470.00	mar24
2,359,712.00	2,203,903.15	2,380,333.93	2,073,373.84	2,138,411.00	2,359,712.00	jun24
2,390,582.00	2,203,903.15	2,408,470.43	2,087,156.70	2,138,411.00	2,390,582.00	sep24

Table 8 presents the comparison of actual and predicted balance sheet values using artificial intelligence models for a Level IV financial institution, for quarterly periods in 2024. The results maintain the trend observed in the previous operational levels, evidencing differences in prediction accuracy among the evaluated models. In particular, the MLP model optimized by genetic algorithms –MLP_GA– presents the closest values to those observed, confirming its better performance in estimating the financial series. This consistent behavior at the highest operational level reinforces the model's ability to generalize in scenarios of greater financial complexity. Likewise, the results are consistent with the evaluation metrics previously reported, which supports the robustness of the proposed approach.

Table 8. Comparison of actual and predicted balance sheet values for a Level IV financial institution in 2024

LR [$]	RF [$]	MLP_GA [$]	MLP [$]	DT [$]	Balance [$]	Period
5,952,100.00	4,859,310.21	5,701,407.64	5,333,042.76	5,477,235.40	5,952,100.00	mar24
6,141,047.00	4,859,310.21	5,813,337.59	5,459,675.43	5,477,235.40	6,141,047.00	jun24
6,322,682.00	4,859,310.21	5,920,573.50	5,577,529.16	5,477,235.40	6,322,682.00	sep24

Figure 14, Figure 15, Figure 16, and Figure 17 present the distribution of the Mean Absolute Error – MAE obtained for all evaluated financial accounts across the four analyzed financial institutions using boxplot representations. Each figure summarizes the predictive errors generated by the evaluated models, allowing a comparative analysis of their central tendency and variability. The results show that the MLP model optimized using genetic algorithms –MLP_GA– consistently presents the lowest MAE values compared to the other evaluated techniques, indicating greater prediction accuracy. In the boxplots, the distribution associated with MLP_GA is concentrated closer to the median and exhibits lower dispersion, suggesting more stable and consistent predictive behavior across the evaluated accounts. In contrast, the remaining models exhibit greater variability and more dispersed error distributions, reflecting lower stability in their predictive performance. This behavior remains consistent across the four financial institutions analyzed. The comparative analysis based on MAE distributions demonstrates that MLP_GA achieves a lower average prediction error while maintaining greater robustness across different financial scenarios and account structures. This consistency across different institutions reinforces the reliability of the proposed model for financial statement forecasting.

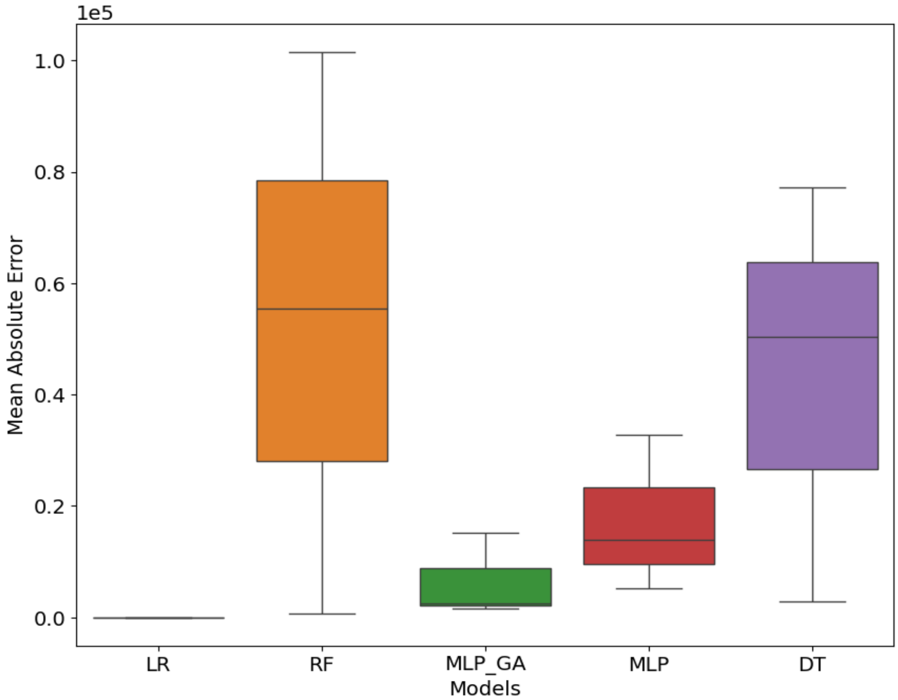

Figure 14. Comparison of prediction errors MAE for Institution I using artificial intelligence models

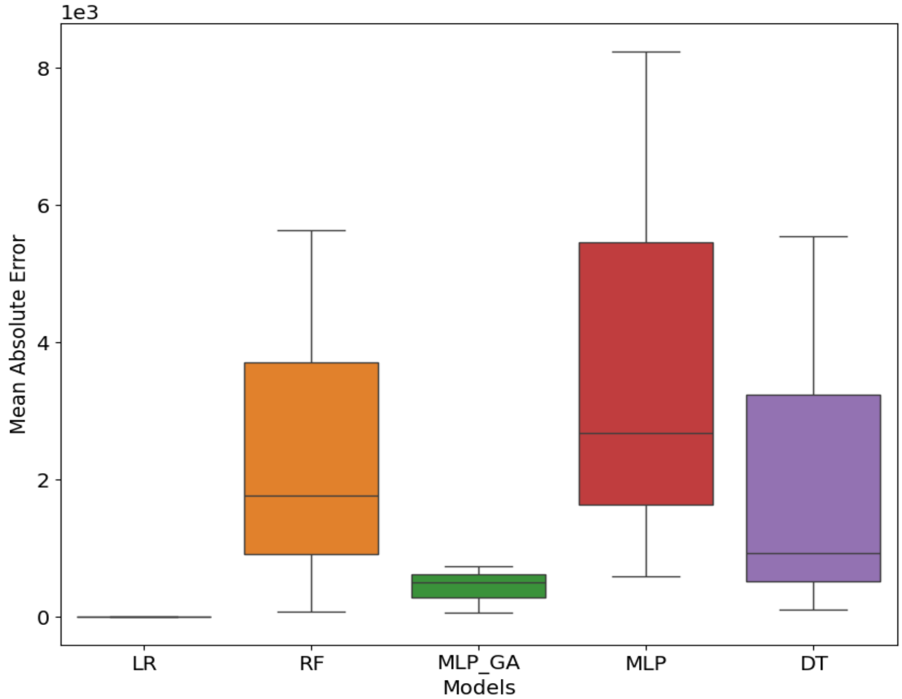

Figure 15. Comparison of prediction errors MAE for Institution II using artificial intelligence models

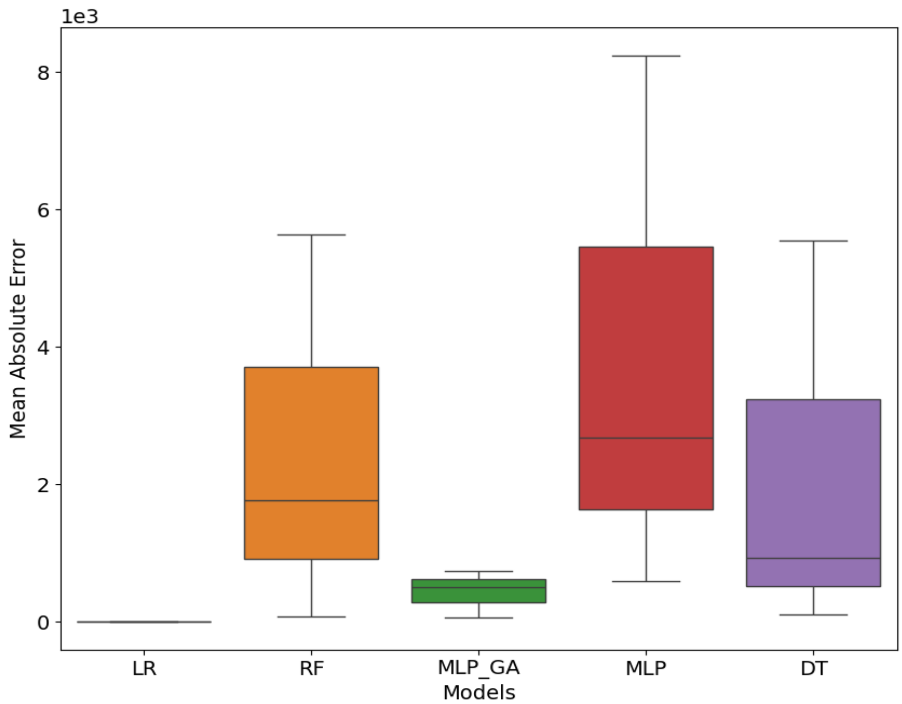

Figure 16. Comparison of prediction errors MAE for Institution III using artificial intelligence models

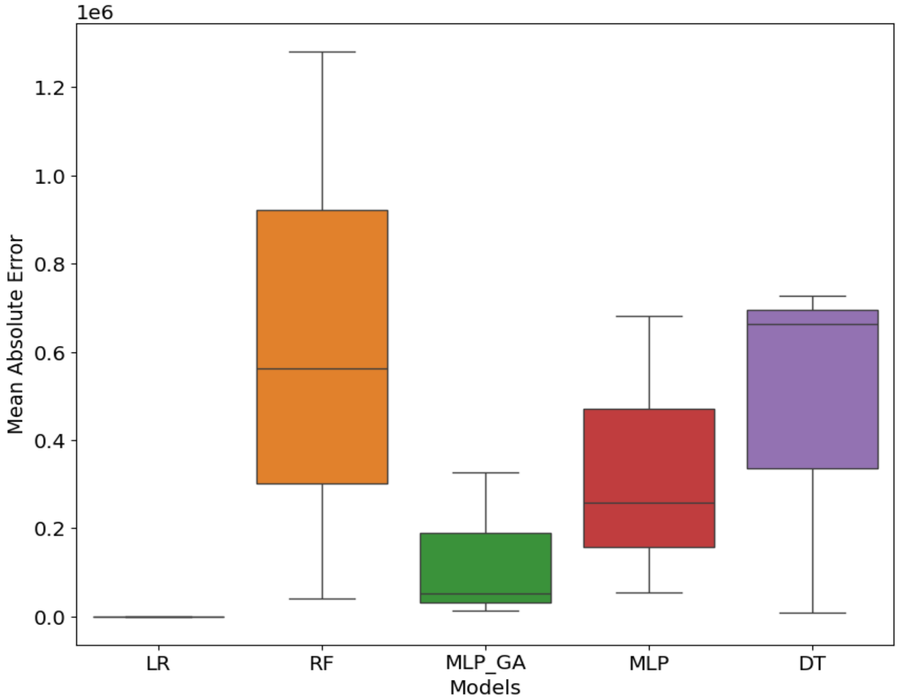

Figure 17. Comparison of prediction errors MAE for Institution IV using artificial intelligence models

The monthly analysis performed in this study showed that the MLP_GA model could capture short-term variations and seasonal fluctuations in financial accounts more effectively than the other evaluated models. In particular, the model demonstrated a greater ability to adapt to abrupt changes and irregular patterns observed in specific periods of the financial series. The use of monthly data provided a finer level of granularity, allowing the model to learn localized temporal behaviors and dynamic variations in the analyzed accounts.

In contrast, the quarterly analysis provided a more aggregated and stable representation of the financial series, facilitating the identification of medium-term trends and reducing the impact of short-term fluctuations. Under this evaluation scheme, the MLP_GA model maintained consistent predictive performance, showing lower variability in the prediction errors compared to the other techniques.

The combined use of monthly and quarterly analyses contributed to a more comprehensive evaluation of the proposed approach. While the monthly analysis allowed the assessment of short-term predictive capability, the quarterly evaluation provided evidence of the model's stability and generalization capacity over broader financial periods. Overall, the MLP_GA model demonstrated greater robustness and consistency in financial forecasting across different temporal resolutions.

The superior and consistent predictive performance achieved by the MLP_GA model suggests that hyperparameter optimization plays a critical role in improving the adaptability of artificial intelligence models to financial time series data. In this regard, genetic algorithms constitute an efficient strategy for exploring complex search spaces in hyperparameter optimization, particularly in forecasting problems characterized by nonlinear relationships and dynamic temporal patterns.

Figure 18 presents the scheme of the genetic algorithm used for hyperparameter optimization in the proposed financial prediction model. In a first stage, the model hyperparameters, such as the number of neurons, activation function, and learning rate, are discretized and encoded as chromosomes, allowing their manipulation within the evolutionary process.

Subsequently, an initial population of solutions is randomly generated to cover different regions of the search space. Each individual is evaluated using a fitness function –see Equation (4)–, defined based on the error metrics employed in the study, which allows quantifying its performance in financial series prediction.

Based on this evaluation, genetic operators of selection, crossover, and mutation are applied to generate new solutions, favoring both exploration and exploitation of the search space. This iterative process is repeated for a predefined number of generations or until a convergence criterion is reached, established based on the minimization of prediction error.

Finally, the selected individual corresponds to the optimal hyperparameter configuration, which is used for training the MLP_GA model, improving its predictive capacity in the context of financial forecasting.

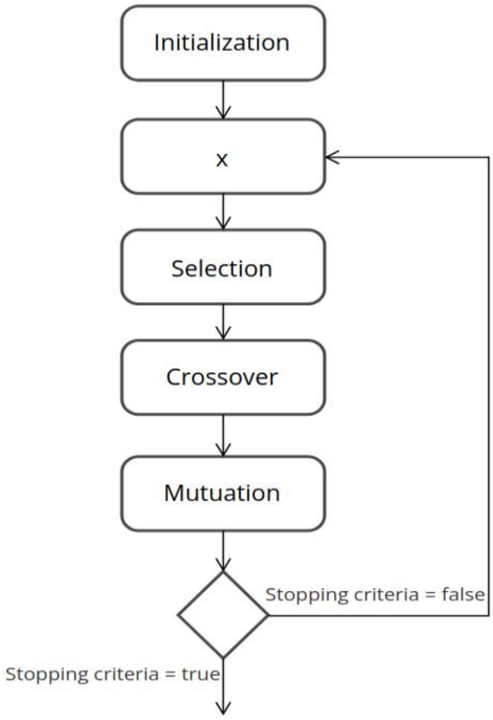

Figure 18. Scheme of the genetic algorithm for hyperparameter optimization

5.1 Statistical tests

To quantitatively validate the differences observed among the evaluated models, the non-parametric Wilcoxon signed-rank test was applied. This test was selected because it does not assume normality in the data and is suitable for comparing related samples, as is the case of the prediction errors obtained by different models over the same time periods.

For sample construction, absolute errors per period were calculated from the actual values and the predictions generated by each model. Periods with null values or no variation were excluded to avoid biases in the analysis. Subsequently, the MLP model optimized by genetic algorithms –MLP_GA– was compared against the LR, MLP, DT, and RF models.

Table 9 presents the results of the Wilcoxon test applied to compare the MLP_GA model with the other evaluated approaches. As observed, in all cases, p-values are less than 0.05, indicating the existence of statistically significant differences between the proposed model and the compared models.

These results confirm that the MLP_GA model presents superior performance in terms of prediction error, quantitatively validating the observations obtained from the descriptive analysis.

Table 9. Wilcoxon test results

Comparison	p-value	Interpretation
MLP_GA vs LR	0.031	Significant difference
MLP_GA vs DT	0.031	Significant difference
MLP_GA vs RF	0.031	Significant difference
MLP_GA vs MLP	0.031	Significant difference

6 Conclusions and future work

In this study, neural networks, specifically the Multilayer Perceptron – MLP, outperformed artificial intelligence techniques such as decision trees, linear regression, and random forests in financial budget prediction under the evaluated conditions. The MLP's ability to capture complex nonlinear relationships in the financial data used allowed it to provide more accurate and reliable predictions within our experimental framework. This budgetary projection performance is attributed, in our models, to the multilayer structure of the MLP, which facilitated the learning of deep and abstract features from the input data.

The implementation of the hybrid MLP_GA method, which combines MLP with genetic algorithms for hyperparameter optimization, showed a significant improvement in budget prediction in our experiments. Genetic algorithms optimized MLP hyperparameters, such as the number of neurons in each layer and the learning rate, resulting in a more efficient and effective neural network configuration for the datasets tested. This hybrid approach not only improved prediction accuracy in our study but also reduced the risk of overfitting compared to a standard MLP, providing a more robust and generalizable model within the scope of our evaluation.

The results obtained in this work highlight the potential of the hybrid MLP_GA method to improve financial budgeting processes and decision-making, at least for the financial data and conditions considered. By leveraging the advanced pattern recognition capabilities of MLP and hyperparameter optimization through genetic algorithms, we achieved superior predictive budgeting in our experiments. These findings suggest that, under similar conditions, the adoption of hybrid artificial intelligence techniques could offer competitive advantages to financial entities, allowing them to make more accurate forecasts and informed data-driven decisions.

The analysis of financial predictions using the hybrid MLP_GA method, based on our dataset, revealed a significant improvement in forecast accuracy and reliability. This enhanced accuracy, observed in our experiments, would allow financial entities to anticipate budgetary variations more precisely, facilitating more effective financial planning and resource management, provided that their data exhibit similar patterns. The ability of MLP_GA to adapt to the different data patterns present in our study and to optimize its parameters based on the specific market conditions analyzed made it a valuable tool for strategic decision-making within the financial sector context we examined.

In the monthly analysis conducted in this study, it was observed that the MLP_GA model was able to capture short-term variations and seasonal fluctuations in financial accounts more effectively than the other evaluated models. In particular, the model showed a greater capacity to adapt to abrupt changes, and irregular patterns present in specific periods of the analyzed financial series. The use of monthly data, in this context, provided a level of granularity that allowed the model to learn localized temporal behaviors and dynamic variations in the accounts under consideration.

In contrast, the quarterly analysis carried out in this study yielded a more aggregated and stable representation of the financial series, which facilitated the identification of medium-term trends and reduced the impact of short-term fluctuations within the study's data. Under this evaluation scheme, the MLP_GA model maintained consistent predictive performance, showing lower variability in prediction errors compared to the other techniques included in the comparison.

The combined use of both analyses –monthly and quarterly– contributed to a more comprehensive evaluation of the proposed approach. While the monthly analysis allowed the assessment of the model's short-term predictive capability within the available time series, the quarterly evaluation provided evidence –limited to the scope of this study– regarding the model's stability and generalization capacity over broader financial periods. Overall, within the experiments performed, the MLP_GA model demonstrated greater robustness and consistency in financial forecasting across different temporal resolutions.

The superior and consistent predictive performance achieved by the MLP_GA model in this work suggests that hyperparameter optimization plays a relevant role in improving the adaptability of artificial intelligence models to financial time series data, at least for the datasets and experimental conditions considered here. In this regard, genetic algorithms constituted, within the context of this study, an efficient strategy for exploring complex search spaces in hyperparameter optimization, particularly in forecasting problems characterized by nonlinear relationships and dynamic temporal patterns such as those analyzed.

For future work, it is recommended to explore the application of other emerging artificial intelligence techniques in financial budget prediction. This includes the use of more advanced deep learning models, such as convolutional neural networks and recurrent neural networks, which could capture temporal and spatial patterns in financial data. Likewise, the integration of natural language processing techniques could enable the analysis of textual data, such as financial reports and economic news, further improving prediction accuracy.

Conflict of interest statement

The authors declare no conflict of interest regarding the research, authorship, or publication of this article.

Acknowledgments

The authors wish to express their gratitude to the Public University of Panama, Coclé Campus, for their valuable support and collaboration in the development of this research.

ORCID ID

Patricia Jimbo Santana https://orcid.org/0000-0001-7432-1622
Erwis Melchor Pérez https://orcid.org/0000-0001-6034-2676
Monica Jimbo Santana https://orcid.org/0000-0002-3948-9507

References

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The journal of Machine Learning Research, 13*, 281-305. https://doi.org/10.5555/2188385.2188395

Chung, H., & Shin, K. S. (2018). Genetic algorithm-optimized long short-term memory network for stock market prediction. *Sustainability, 10*(10), 1-18, 3765. https://doi.org/10.3390/su10103765

De Rainville, F. M., Fortin, F. A., Gardner, M. A., Parizeau, M., & Gagné, C. (2012). DEAP: A Python framework for evolutionary algorithms. *Companion proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2012)* (pp. 85-92). Association for Computing Machinery. https://doi.org/10.1145/2330784.2330799

Dietterich, T. G. (2000). *Ensemble methods in machine learning* (Vol. 1857). MCS 2000. https://doi.org/10.1007/3-540-45014-9_1

Gülmez, B. (2025). GA-Attention-Fuzzy-Stock-Net: An optimized neuro-fuzzy system for stock market price prediction with genetic algorithm and attention mechanism. *Heliyon, 11*(3), e42393. https://doi.org/10.1016/j.heliyon.2025.e42393

Gür, Y. E., Yıldız, A., & Ünal, E. (2025). Advanced AI models for future forecasting of budget expenditures via machine learning and deep learning. *Panoeconomicus*, 1-36. https://doi.org/10.2298/PAN240929025G

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial Intelligence and Machine Learning in Finance: Identifying Foundations, Themes, and Research Clusters from Bibliometric Analysis. *Journal of Behavioral and Experimental Finance, 32*, 100577. https://doi.org/10.1016/j.jbef.2021.100577

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press. http://www.deeplearningbook.org

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), pp. 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice* (3rd ed.). Melbourne, Australia: OTexts. https://otexts.com/fpp3/

Khattak, B. H., Shafi, I., Khan, A. S., Flores, E. S., Lara, R. G., Samad, M. A., & Ashraf, I. (2023). A systematic survey of AI models in financial market forecasting for profitability analysis. *IEEE Access, 11*, pp. 125359-125380. https://doi.org/10.1109/ACCESS.2023.3330156

Li, W., & Law, K. E. (2024). Deep learning models for time series forecasting: A review. *IEEE Access, 12*, 92306-92327. https://doi.org/10.1109/ACCESS.2024.3422528

Lin, S., Qiao, Z., Chen, J., & Huang, Z. (2025). A hypertension early warning model combining generative adversarial networks and long short-term memory neural networks. In *Web Information systems Engineering - WISE 2024* (Vol. 15463, pp. 341-356). Springer. https://doi.org/10.1007/978-981-96-1483-7_29

Liu, S. (2025). LTR-Net: A deep learning-based approach for financial data prediction and risk evaluation in enterprises. (U. o.-E. Rahul Gomes, Ed.) *PLOS One, 20*(8), 1-20, e0328013. https://doi.org/10.1371/journal.pone.0328013

Mandal Das, R., Sarkar, B., & Sarkar, A. (2025). A study on stock market analysis through artificial intelligence. https://doi.org/10.2139/ssrn.6468375

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005-2019. *Applied Soft Computing Journal, 90*, 106181. https://doi.org/10.1016/j.asoc.2020.106181

Shetty, S., Musa, M., & Brédart, X. (2022). Bankruptcy prediction using machine learning techniques. *Journal of Risk and Financial Management, 15*(1), 35. https://doi.org/10.3390/jrfm15010035

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems, 25*, pp. 2951-2959. https://doi.org/10.48550/arXiv.1206.2944

Sun, L., Diao, S., Lyu, B., & Chen, K. (2025). Integration of genetic algorithms and trading simulation for enhanced stock price prediction and optimization of quantitative trading strategies. *Proceedings of the 2024 2nd International Conference on Electronics, Computers and Communication Technology* (pp. 166-171). New York: ACM. https://doi.org/10.1145/3705754.3705787

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *31st International Conference on Neural Information Processing Systems (NeurIPS 2017)*, *30*, pp. 5998-6008. https://neurips.cc/Conferences/2017

Wilson, G. T. (2016). Book Review: Time Series Analysis: Forecasting and Control, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1. *Journal of Time Series Analysis, 37*(5), 709-711. https://doi.org/10.1111/jtsa.12194

Zhang, S., Hu, X., Gao, Y., Gao, M., & Zhang, Y. (2025). Exemplar learning and memory retrieval-based particle swarm optimization algorithm with engineering applications. *Biomimetics, 10*(10), 708. https://doi.org/10.3390/biomimetics10100708