Texto para Discussão 006 | 2023

Discussion Paper 006 | 2023

Redes Neurais Artificiais para Análise e Previsão de Fenômenos Econômicos: Avaliando o Grau de Monetização da Economia Brasileira

Kleyton Vieira Sales da Costa
Economista pela Universidade Federal Rural do Rio de Janeiro
Pesquisador na Holistic AI

Antonio Jose Alves Junior
Professor Titular do Departamento de Economia da Universidade Federal Rural do Rio de Janeiro

Felipe Leite Coelho da Silva
Professor Associado do Departamento de Matemática da Universidade Federal Rural do Rio de Janeiro
Pesquisador do CNPq

André de Melo Modenesi
Professor Associado do Instituto de Economia da Universidade Federal do Rio de Janeiro Pesquisador do CNPq e da FAPERJ

Débora Mesquita Pimentel
Professora Associada do Departamento de Economia da Universidade Federal Rural do Rio de Janeiro

Redes Neurais Artificiais para Análise e Previsão de Fenômenos Econômicos: Avaliando o Grau de Monetização da Economia Brasileira

Março, 2023

Kleyton Vieira Sales da Costa
Economista pela Universidade Federal Rural do Rio de Janeiro
Pesquisador na Holistic AI
kleyton.vsc@gmail.com

Antonio Jose Alves Junior
Professor Titular do Departamento de Economia da Universidade Federal Rural do Rio de Janeiro
antonio.alves.2005@gmail.com

Felipe Leite Coelho da Silva
Professor Associado do Departamento de Matemática da Universidade Federal Rural do Rio de Janeiro
Pesquisador do CNPq
felipeleiterural@gmail.com

André de Melo Modenesi
Professor Associado do Instituto de Economia da Universidade Federal do Rio de Janeiro Pesquisador do CNPq e da FAPERJ
amodenesi@gmail.com

Débora Mesquita Pimentel
Professora Associada do Departamento de Economia da Universidade Federal Rural do Rio de Janeiro
pimentel.deboram@gmail.com

Variable Selection of Economics Phenomena with Explainable Artificial Intelligence Methods*

Kleyton da Costa†[1], Antonio Jose Alves Junior[2], Felipe Leite Coelho da Silva[3], Andre de Melo Modenesi[4], and Débora Mesquista Pimentel[2]

[1]Departament of Informatics, PUC-Rio, Brazil

[2]Departament of Economics, UFRRJ, Brazil

[3]Departament of Mathematics, UFRRJ, Brazil

[4]Institute of Economics, UFRJ, Brazil

March 4, 2023

Abstract

This study aims to evaluate the ability of explainable artificial intelligence (XAI) models in the selection of explanatory variables for economics phenomena. A workflow was built considering machine learning models (Random Forest and Multilayer Perceptron Regressor) and conventional multiple regression models (Generalized Linear Model, Bayesian Regression, Robust Regression, and Ridge Regression). The dependent economic variable selected was the degree of monetization for Brazilian economy in period between January 2001 and January 2021. The behavior of the degree of monetization can be impacted by the risk perceived by economic agents concerning the future behavior of the economy and, therefore, considering this factor, variables usually used in the forecasts of economic agents were selected, namely: Selic rate, GDP growth rate, exchange rate, and inflation rate. The results found by measuring the SHAP value indicate that the most important variables to predict the degree of monetization of the economy are, respectively: the Selic rate, the exchange rate, and the inflation rate. Finally, a comparative study was applied between artificial neural network models (Recurrent Neural Network, Long Short-term Memory, and Gated Recurrent Unit) to predict the degree of monetization considering the variables selected by the SHAP method as the input values. The study showed that the Recurrent Neural Network model has the best predictive capacity among the selected models.

*Artigo apresentado no 50º Encontro da ANPEC, 2022

†kleyton.vsc@gmail.com

1 Introduction

The degree of monetization of the economy is represented in this study by the ratio between the total currency held by the public and demand deposits and the total currency invested in interest-earning assets. To calculate the degree of monetization of the economy, two components that belong to the so-called Monetary Aggregates (AM) are used: restricted means of payment and financial savings. The AM are historical series that represent an aggregation of different forms of currency for different levels of liquidity, usually in descending order from the most liquid assets to the least liquid assets[1] [30].

The idea of a hierarchy established based on the degree of liquidity of an economy's assets is present in [14] through the concepts of fully liquid, liquid, and illiquid assets. In addition to currency itself, fully liquid assets are those that can be converted into currency without significant transaction costs and without delay. Liquid assets are those traded on organized markets and with well-defined trading rules, such as scheduled times and dates for exchanges to take place. Agents usually acquire liquid assets with the expectation of their future appreciation, even if there is some carrying cost, as is the case of works of art and government bonds. Illiquid assets are those whose trading takes place in unorganized markets, with the resale price lower than the acquisition price, such as consumer durables held by individuals or capital equipment held by companies. However, the threshold between the degree of liquidity of assets may be less clear due to the way the market is organized and the consequent characteristics inherent to this market.

Monetary authorities, in the form of central banks, are responsible for issuing currency in countries that have their own currency. However, as described in [7, p.19], only the portion of the issued resources that are recorded as bank deposits and paper money held by the public is considered monetary issues. The portion that remains on the central bank's cash is not legally considered currency. The composition of the restricted means of payment (M1) occurs through the sum between the volume of paper money held by the public and demand deposits. These are considered fully liquid assets, taking into account that they fully perform the function of store of value, being able to settle debts formed through formal contracts or obligations from transactions in the spot market [7].

With the growth of computational capacity, machine learning models have become popular for applications in various fields of knowledge, including economics. A machine learning model is a program that maps inputs and generates outputs. As they are "black box" models, there are barriers to their use, given the inability to meet transparency and auditing requirements. Broadly speaking, an explainable or interpretable algorithm is one in which the reasons for a decision of some "black box" model can be questioned and explained in a way that makes sense to humans. Thus, *explainability* is the ability

[1]Note that the modern monetary and financial system has financial instruments with high liquidity and interest income, such as savings accounts.

to explain what is happening in the algorithm in terms of variables more or less relevant to the prediction, and *interpretability* is the ability to convey these results in humanly understandable terms. [10].

In this study, we used an explainability model to measure the individual (and average) impact for a set of variables in relation to the prediction of a target variable. Thus, through the results generated by the models, it is possible to identify which are the most relevant variables. Taking into account that the degree of monetization of an economy is a variable of interest to investigate the behavior of economic agents in relation to restricted means of payment and the economy's financial savings, this chapter estimates the impact of macroeconomic variables on the degree of monetization. of the Brazilian economy for the period from December 2001 to January 2021 and performs a forecasting exercise using deep learning models.

The article is structured as follows: Section 2 a brief literature review, Section 3 presents the methodology used, Section 4 describes and discusses the main results found and, finally, Section 5 presents the conclusion and proposals for future research.

2 Literature Review

The construction of a workflow proposed by [6] takes into account the *interpretability* of machine learning models as an efficient instrument for predicting economic phenomena. The study analyzed US employment data using a set of machine learning models and conventional linear models, concluding that the former performed better. In addition, *Shapley* values were applied to identify nonlinear behaviors that have economic significance.

The study proposed by [1] uses SHAP values to select variables that can predict inflation in an emerging economy (Turkey) more accurately. The study proposed the comparison between factor models and machine learning models, resulting in a better performance of machine learning models.

Time series forecasting has methodological approaches that are presented through different classes of models and methods. An univariate approach can be built through classical models, widely applied in forecasting economic series. The best known of this class are the Box & Jenkins [4] models, which is an expanded form that takes into account the seasonality that can be identified in a series through the autocorrelation function. The exponential smoothing models [[17], [34]] are also methodological approaches with different applications to real data, taking level, trend, and seasonal smoothing equations that can significantly adjust to quarterly data. The class of state-space models [33] is another that has great relevance and applicability to economic and financial data. An exponent of this class is the dynamic linear models that, through a robust and sophisticated Bayesian framework, can generate expressive results, as proposed in [9]. In addition, artificial neural network models gain more and more space for time series analysis and prediction.

Applications of neural networks for forecasting exchange rates have results that are influenced by problems related to data selection and sample variation. Most applications adopt a model validation practice that consists of separating the data between a training set (*in-sample*) and a test set (*out-sample*). [18] notes two main problems with this way of selecting data for training and testing. Bias is introduced in the model when the observed characteristics of the training set are different from those observed in the test set. And the effects that the size of the training set has on the model are not taken into account. Thus, the author proposes to use a *cross-validation* methodology to analyze the *out-sample* performance of neural networks. The study uses applied neural networks to predict the weekly exchange rate data between the GBP and USD currencies. The results obtained indicate that the accuracy of neural networks does not have a significant sensitivity concerning sample variation. Another result found is that the neural network model makes predictions with better performance for the short term, also obtaining more significant results than those observed in the *random walk*.

The study proposed by [37] examines how the number of hidden layers, input layers and sample size impact the performance of in-sample and out-of-sample predictions. The exchange rate analyzed is between GBP/USD, with daily data covering the period between 1976 and 1993. The model used was the commonly used *multilayer perceptron*. As a way of stabilizing the series, the authors chose to use a logarithmic transformation. The study points out that neural networks have more significant results than linear models, especially in the short term. The number of inputs has as much impact on model performance as the number of hidden layers, as a greater number of observations contributes to the reduction of prediction errors. Predictions were generated weekly, pointing out that the selection of neural network inputs and architecture are significant factors for predictive effectiveness.

An application for forecasting the exchange rate was proposed by [36]. Authors use an approach that takes into account the "union of neural network models" (*neural networks ensembles*) to improve the performance of the neural network, being a different approach from the standard idea of choosing a single best model. Thus, the article presents the union of neural networks for the forecast of the exchange rate. The results show that the combination of neural networks generated a significant improvement in results compared to individual neural networks. But on the other hand, the results did not get significantly better results than the *random walk* model.

The methodology proposed in [35] proposes a non-linear combination between a generalized autoregressive linear model (GLAR) and an artificial neural networks (ANN) model, seeking to increase the predictive accuracy. The hybrid model was compared with its isolated versions and applied to data on the exchange rates DEM/USD, GBP/USD and JPY/USD in monthly values, for the period between January 1971 and December 2000. The results indicate that the proposed hybrid model (GLAR-ANN) obtained a better prediction result than the GLAR and ANN models.

The comparative study proposed by [26] investigates a neural network model, an autoregressive linear model and a *random walk*. The justification for using artificial neural networks for the purpose of forecasting the exchange rate between currencies is the non-linearity observed in this type of series, causing the linear models to have unsatisfactory results. The work uses the weekly series of the exchange rate between INR (*Indian Rupee*)/USD, covering the period between January 6, 1994 and July 10, 2004 and proposing a one-step forecast. The results found indicate that the neural network model is superior to the linear autoregressive model and *random walk* for the prediction within the sample. For the out-of-sample prediction, the neural network model was partially superior, obtaining a better performance for most of the considered metrics.

The literature is also composed of studies that use artificial neural network models to forecast financial assets. [32] investigates the use of a combination of neural network models in order to improve the predictive capacity of the models. For the empirical analysis, credit data from Australia and Germany were used, in addition to bankruptcy data provided by the *Standard and Poor's* rating agency. The results found indicate that the combination of models performed better than the strategy of a single best model. Another application of hybrid/combination models was developed in [11]. In this work, the methodology used consists of a hybrid model of financial systems (HFSs) combining artificial neural networks with ARIMA or ARCH/GARCH models. For the analysis, data from the Karashi stock exchange index (KSE100) were used, in daily values, for the period between January 1, 2000 and October 18, 2002. The hybrid model ($ANN_{ARCH/GARCH}$) presented the best performance compared to the other models considered in the study (ANN, ARIMA and GARCH).

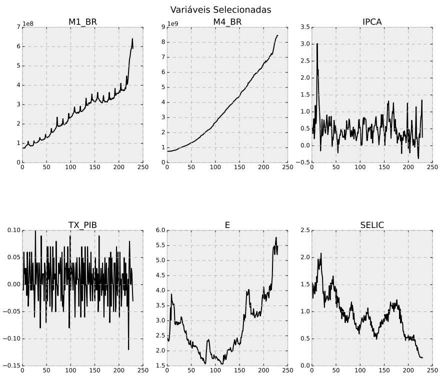
3 Methodology

This section presents the methodology used in this work. Two modeling approaches were implemented: explanatory models and predictive models for the degree of monetization of the Brazilian economy. Explanatory models are intended to determine which variables are significant to explain the degree of monetization. In addition, an interpretability model was applied to identify which variables have the greatest impact on the results of the models.

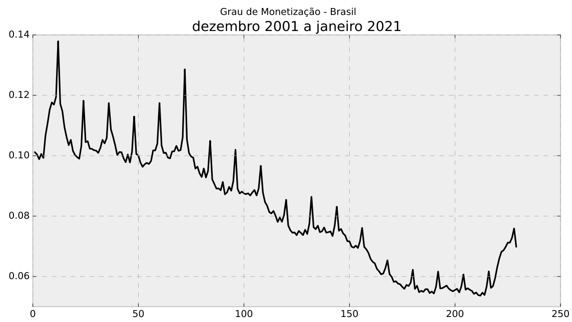
After validating the selected variables, artificial neural network models were implemented for time series prediction. At this stage, a simulation exercise was carried out using different configurations for the initial parameters. The performance of the models was compared using a conventional approach through precision metrics.

3.1 Data Description

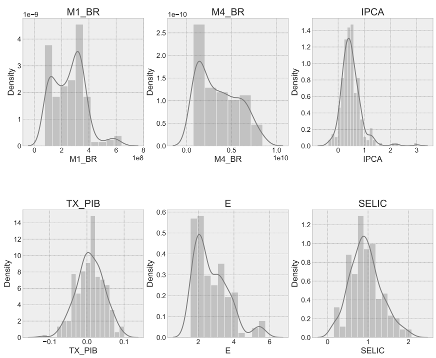
To conduct the experiments, the variables listed below were selected, described in Table 1 and presented in Figure 1. All data were obtained through the BCB Time Series Management System [2]. It is observed (Table **??**) that the variables have 229 observations, covering the period between January 2001 and January 2021. The selected variables are: M1 - Restricted means of payment (balance at the end of the period) - u.m.c. (thousand); M4 - Financial savings. Large means of payment - (balance at end of period) - c.u. (thousand); IPCA - Broad consumer price index; SELIC - Interest rate, accumulated Selic in the month; GDP - Monthly gross domestic product growth rate at current values; E - Exchange rate for the Brazilian real against the US dollar.



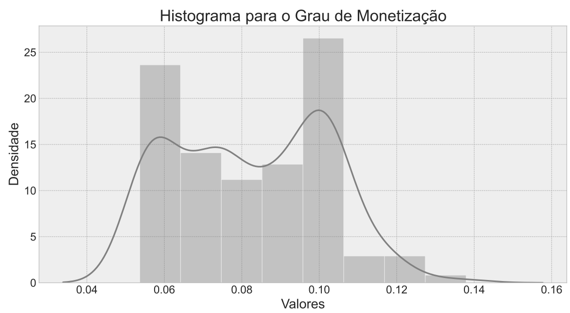
(a) Variables



(b) Degree of monetization



(c) Variable histogram



(d) Degree of monetization histogram

Figure 1: Variables and degree of monetization for Brazilian economy from 2001, December to 2021, January

	M1	M4	IPCA	PIB	E	SELIC
# Observations	229	229	229	229	229	229
Mean	254412701.2620	3560907502.4934	0.4925	0.0083	2.7555	0.9401
Std. Deviation	118050957.9182	2221602631.0116	0.3897	0.0370	0.9576	0.3858
Min	74996003.0000	753801974.0000	-0.3800	-0.1153	1.5600	0.1500
25%	147551340.0000	1449901658.0000	0.2600	-0.0147	2.0200	0.6900
50%	262523711.0000	3236156038.0000	0.4400	0.0070	2.4200	0.9100
75%	326435805.0000	5419606850.0000	0.6400	0.0339	3.2800	1.1600
Max	641107547.0000	8454294566.0000	3.0200	0.1015	5.7700	2.0800

Table 1: Summary of variables

3.2 Regression Models

To investigate the elements that affect the degree of monetization of the Brazilian economy and the sign of these effects, a regression strategy was used in order to measure the level of explainability that macroeconomic variables exert on the analyzed phenomenon. The proposed explanatory models were built from six approaches: (i) generalized linear model (Gaussian); (ii) generalized linear model (Gamma); (iii) Bayesian regression; (iv) regression *Ridge*; (v) robust regression and; (iii) *Multilayer Perceptron Regressor*.

To measure the impact (and the average impact) of the selected variables on the explanatory capacity of the proposed models, the SHAP [22] model was used, a computational strategy based on the *Shapley value*. The use of Shapley values to identify which variables are most relevant to explain an economic phenomenon from machine learning methods is an alternative to traditional econometric methods [6].

Generalized Linear Regression An extension of the simple linear models proposed by [25] was considered for application to the class of generalized linear models. The model considers a combination from a systematic component and a random component. Thus, the extension from a conventional linear model occurs through a link function (θ) applied through a linear combination between the predictors ($\hat{y} = \sum_{i=1}^{m} \beta_i x$) and the set of observations $x_i = (1, \ldots, i)$. This process can be expressed as

$$\hat{y}_i(\beta_i, x_i) = \theta(x_i \beta_i) \tag{1}$$

where β_i are the coefficients associated with the variables, with $\beta_i = (1, 2, \ldots, i)$. The link function $\theta = (x_i \beta_i)$ connects the parameter θ of the distribution z with the values observed in the linear regression. For the study, a Gaussian distribution - $\sum(z - \mu^2)/\sigma^2$ was considered. Using the Gaussian distribution we obtain a linear regression model with errors that follow $N(\theta, \sigma^2)$.

Regularized Ridge Regression Regression models that use regularization are strategies to deal, in essence, with overfitting problems, i.e., when models can significantly adjust to observed data but do not make good predictions, considering a given set of variables. train $x_i = (1, \ldots, n))$ and test $y_i = (n+1, \ldots, m)$. In this study, the regression approach with regularization *Ridge* (L_2) was applied. A penalty like L_2 adopts a restriction on the coefficients through a penalty factor. The regularization strategy considered in the L_2 method uses the squares of the coefficients [16] estimated by the regression model, that is,

$$L_2(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2 + \lambda \sum_{j=1}^{m} \hat{\beta_j}^2 \tag{2}$$

where λ determines the size of the constraint that will be imposed on the regression. The greater the value of λ, the greater the penalty of the optimization function, considering that $\lambda = 0$ represents a simple linear model. In addition to reducing model complexity (and overfitting), L_1 regularization reduces multicollinearity.

Bayesian Regression Bayesian regression is used to add regularization parameters in the estimation process through an iteration process. This Bayesian approach to linear regression circumvents the overfitting problem observed through maximum likelihood estimation [3]. A Bayesian approach to *Ridge* regression is constructed by minimizing $\hat{\beta}$ as the average *lateri* of a model in which

$$\beta_i \sim N(0, \sigma^2/\lambda), \forall i. \tag{3}$$

Robust Regression The robust regression [27] aims to fit a regression on data that have a non-standard behavior, ie, have *ouliers* or specification errors. In this study, a Huber estimation strategy was used (H_ϵ), applying a linear error function to the sample data that are identified as *outliers*. Determining whether a sample is not an unusual observation occurs if the absolute error of the sample is less than a *threshold*. In this way, the error function for a Huber estimator given by,

$$\min_{w,\sigma} \sum_{i=1}^{n} \left(\sigma + H_\epsilon \left(\frac{X_i w - y_i}{\sigma} \right) \sigma \right) + \alpha ||w||_2{}^2 \tag{4}$$

considering the estimator function as,

$$H_\epsilon(z) = \begin{cases} z^2, \text{ if } |z| < \epsilon \\ 2\epsilon|z| - \epsilon^2, \text{ otherwise.} \end{cases} \tag{5}$$

Random Forest A Random Forest is characterized as an algorithm for combining a set of decision trees in which each tree depends on the values of a random vector with independent samples and with the same [5] distribution. The decisive modification with the original method (Bagging) is in the development of an expressive number of decision trees that are not correlated and calculating their average from that point. The 1 Algorithm presents the steps that must be followed for a Random Forest to be applied for regression purposes.

Algorithm 1: Random Forest Algorithm [13]

---

1. considerando $b = (1, \ldots, B)$

   (i) Considerar uma amostra *bootstrap* Z^* e com um conjunto de treino de tamanho X

   (ii) Definir uma *random forest tree* T_b para os dados *bootstrap*, repetindo recursivamente as seguintes etapas para cada nó terminal da árvore, até que o tamanho mínimo do nó seja alcançado.

   (a) *selecionar m variáveis de forma aleatória de p variáveis*

   (b) *selecionar a melhor variável entre as m*

   (a) *dividir o nó entre outros dois nós*

2. saída com a combinação de árvores $T_b^B{}_1$

A previsão de um ponto x considerando uma regressão é: $\hat{f}_{rf}^B(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$

Multilayer Perceptron Regressor A Multilayer Perceptron Regressor (MLPR) model performs the training process through backpropagation without applying an activation function in the output layer. In the formulation used in this study, a mean squared error loss function was considered. Furthermore, the outputs are defined from a set of continuous values. The model definition is as follows: given a training set $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$ where $x_i \in \mathbb{R}^n$ and $y_i \in \{0.1\}$, an MLPR model with a hidden layer performs the process learning through the function

$$f(x) = W_{2g}(W_1^T x + b_1) + b_2) \tag{6}$$

where $W_1 \in \mathbb{R}^m$ and the model parameters are $W_2, b1, b2 \in \mathbb{R}$. W_1 and W_2 are the input layer and hidden layer weights, respectively. $b1$ and $b2$ are the hidden layer and output layer biases, respectively. The function $g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is an activation function, in this study, of the hyperbolic tan type which is given by,

$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{7}$$

Finally, the mean square error was used as the loss function ($\mathcal{L}$). So, the function was defined as

$$\mathcal{L}(\hat{y}, y, W) = \frac{1}{2n}\sum_{i=0}^{n} ||\hat{y}_i - y_i||_2^2 + \frac{\alpha}{2n}||W||_2^2 \tag{8}$$

where from randomly defined initial weights, the model minimizes $\mathcal{L}$ from repeated updates to the weights. With the calculation of $\mathcal{L}$, each weight parameter is updated in order to decrease the loss.

SHAP - SHapley Additive exPlanations The definition of the impact of the variables of interest used as input to a conventional or machine learning model can be achieved through methods of explainability. The purpose of the SHAP method is to explain the prediction of a dependent variable from the contribution of each selected independent variable. The innovation that the method presents in relation to the Shapley value $(\phi_i(v))$[2] is the representation given for the value from a described linear model through an additive method for the variables. The method can be described as,

$$\gamma(z') = \phi_0 + \sum_{i=1}^{C} \phi_i(v) z_{i'} \tag{9}$$

where γ is the explanation model, $z' \in \{0,1\}^C$ is the coalition vector, C is the maximum coalition size, and $\phi_i(v) \in \mathbb{R}$ is Shapley value.

Taking as input a set function $v : 2^n \to \mathbb{R}$, we can define $\phi_i(v)$ for a specific variable i [[31], [24], [21]] as your contribution to the payment through the weighted average of all possible combinations

$$\phi_i(v) = \sum_{F \subseteq \{x_1,\ldots,x_n\} \setminus \{x_i\}} \frac{|F|!\,(p - |F| - 1)!}{n!} \left(v\left(F \cup \{x_i\}\right) - v(F)\right) \tag{10}$$

where n is the number of variables, F is the subset of the variables used in the model, and x is the vector of values for the variables in the sample to be explained.

3.3 Forecast Models

To analyze the prediction capacity of artificial neural network models for the degree of monetization of the Brazilian economy, the models *Recurrent Neural Network* (RNN), *Long Short-Term Memory* (LSTM) and *Gated Recurrent Unit* (GRU). Artificial neural network models emerge as an attempt to approximate the functioning of a biological brain through a mathematical function[3]. In this type of modeling, three layers of operation are usually considered: the input layer, the hidden layer, and the output layer.

[2] The Shapley value [29] has become the main contribution to cooperative game theory by allocating *payments* to *players* depending on their contribution to the total. When considering a Machine Learning context, each variable is a *player* that participates in the game; the forecast is the *reward* and the Shapley value communicates how the contribution of each variable (*payments*) can be spread across the variables.

[3] See the seminal discussion of [28] and [23].

Recurrent Neural Network *recurrent neural networks* (RNNs) have been gaining more attention through the increased computational power of graphics processing units (GPUs) [20]. They are useful for time series data since each neuron can use its internal "memory" to hold the information from the previous input. The fundamental problem in a simple RNN is the inability to capture the long-term dependencies in a series, i.e. the recurring connections allow a "memory" of previous inputs to persist in the internal state of the network, influencing the output of the network [12]. We can consider that the *forward propagation* in an RNN is similar to what is observed in an MLP with a single hidden layer, however, the activation functions reach the hidden layer from the current external input, and the activation of the hidden layer from the range of previous time. The model notation considering i the number of input units, H hidden inputs and K output units; x_i^t the value of the input i at the moment t; the network input a_h^t and the final activation b_h^t, both for the unit h at the moment t. The hidden units are given by,

$$a_h^t = \sum_{i=1}^{I} w_{ij} x_i^t + \sum_{h'=1}^{H} w_{h'h} b_{h'}^{t-1}. \tag{11}$$

Long Short-term Memory The LSTM model is a special type of *Recurrent Neural Network* (RNN), developed by [15]. As noted earlier, RNNs have a useful architecture for data organized in time series, since each node uses its internal "memory" to preserve the information that belongs to the previous entry. However, one of the most relevant problems for a simple RNN is the *disappearance gradient problem*, which is an inability to capture the long-term dependence observed in the time series, that is, the recurrent connections allow a " memory" from previous entries persists in the network's internal state, influencing the network output [12]. Thus, LSTM is a solution strategy for this problem, since it considers the long-term dependencies that are present in the time series.

As described in [**?**], for each element in the input sequence, each layer will compute the equations

$$\begin{aligned}
i_t &= \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \\
f_t &= \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) \\
g_t &= \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}) \\
o_t &= \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned} \tag{12}$$

where h_t is the hidden state for time t, c_t is the state cell for time t, x_t is the input for time t, h_{t-1} is the hidden state of the layer at time $t-1$ or the initial state for time 0, and i_t, f_t, g_t, o_t are the input *gates*, oblivion, cell and exit. σ is the activation function (sigmoid), and $\odot$ is the Hadamard product. Considering an LSTM with multiple layers, the input $x_t^{(l)}$ for the l-th layer ($l >= 2$) is the hidden state $h_t^{(l-1)}$ from the previous layer multiplied by *dropout* $\delta_t^{(l-1)}$, where each $\delta_t^{(l-1)}$ is a Bernoulli random variable where 0 is the probability of the *dropout*.

Gated Recurrent Unit Proposed by [8], the GRU model has only two *gates*: a redefining *gate* and an updating *gate*. Like an LSTM model, a GRU seeks to solve the *gradient vanishing problem*. This occurs through two vectors (*gates*) that decide which information will be taken to the output and which will not. In this way, the structure of the model allows capturing the dependence of a considerable volume of data without discarding the previous information of the model, and this in an adaptive way. For each element of the input sequence, each layer will compute

$$
\begin{aligned}
r_t &= \sigma(W_{ir}x_t + b_{ir} + W_{hr}h_{(t-1)} + b_{hr}) \\
z_t &= \sigma(W_{iz}x_t + b_{iz} + W_{hz}h_{(t-1)} + b_{hz}) \\
n_t &= \tanh(W_{in}x_t + b_{in} + r_t * (W_{hn}h_{(t-1)} + b_{hn})) \\
h_t &= (1 - z_t) \odot n_t + z_t \odot h_{(t-1)}
\end{aligned}
\tag{13}
$$

where h_t is the layer's time state at t, x_t is the time input t, $h_{(t-1)}$ is the hidden state of the layer at time $t-1$ or the initial hidden state at time 0, and r_t, z_t, n_t are the restart, update and new *gates gates*. σ is the activation function (sigmoid), and $\odot$ is the Hadamard product. In a multi-tier GRU, the input $x_t^{(l)}$ of the lth tier ($l >= 2$) is the hidden state $h_t^{(l-1)}$ of the previous layer multiplied by *dropout* $\delta_t^{(l-1)}$ where $\delta_t^{(l-1)}$ is a Bernoulli random variable where 0 is the probability of *dropout*.

4 Results and Discussion

This section presents the discussion and the main results found in this study. The first topic to be analyzed concerns the multiple regression estimated considering the degree of monetization of the economy as a dependent variable and having as explanatory variables the growth rate of domestic product monthly product for Brazil, the monthly inflation rate captured through the price index consumer price (IPCA), the exchange rate for the Brazilian real against the US dollar and the basic interest rate (Selic). After the regression analysis, the predictive performance for the considered deep learning models is presented.

4.1 Model Estimation

The composition of financial savings (M4) of an economy is formed by the sum of highly liquid government bonds and expanded means of payment. In turn, the expanded means of payment are determined through the sum of the restricted means of payment and special interest-bearing deposits, savings deposits, securities issued by depository institutions, shares in fixed income funds and registered repo operations. at Selic.

The degree of monetization of the economy can be influenced through a movement of search for security on the part of economic agents, that is, when there is an increase in risk related to the future behavior of the economy, economic agents seek to protect their capital through assets more or less less liquid, depending on the characteristic of the phenomenon that will impact the expected economic behavior. In this way, Table 2 summarizes the expected effect of certain variables in relation to M1 and M4 and, consequently, on the degree of monetization of the economy.

Expected Behavior
When there is an *increase in inflation*, it is assumed that agents tend to migrate from more liquid assets (which do not earn interest) to less liquid assets (which earn interest). Thus, there is a reduction in restricted means of payment and an expansion of financial savings, causing the expected effect of a reduction in the degree of monetization of the economy.
When there is an *increase in the exchange rate*, it is assumed that agents tend to migrate from more liquid assets (which do not earn interest) to less liquid assets (which earn interest). Thus, there is a reduction in restricted means of payment and an expansion of financial savings, causing the expected effect of a reduction in the degree of monetization of the economy.
When there is an *increase in the gross domestic product growth rate*, it is assumed that agents tend to migrate from less liquid assets (which earn interest) to more liquid assets (which do not earn interest). Thus, there is an expansion in restricted means of payment and a reduction in financial savings, causing the expected effect of a reduction in the degree of monetization of the economy.
When there is an *increase in the Selic interest rate*, it is assumed that agents tend to migrate from more liquid assets (which do not earn interest) to less liquid assets (which earn interest). Thus, there is a reduction in restricted means of payment and an expansion of financial savings, causing the expected effect of a reduction in the degree of monetization of the economy.

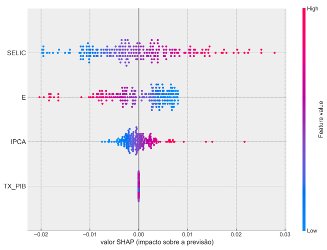
Table 2: Expected behavior of the degree of monetization based on the dynamics of macroeconomic variables.

To measure the degree of monetization of the Brazilian economy, four independent variables estimated using the ordinary least squares method were used. The multiple determination coefficient (R^2) shows consistent results between the different regression strategies adopted. Table 3 shows that, except for the *Random Forest* method, the observed result for the coefficient of determination varies between 0.51 and 0.55, that is, a reasonable result considering the use of a parsimonious number of explanatory variables. Furthermore, through the mean percentage absolute error (MAPE) applied to the model fit result, it shows that the results have a small difference in absolute values, ranging between 0.1212 and 0.1501. The two metrics show that the *Random Forest* method has a better performance, 0.94 for R^2 and 0.0303 for MAPE. Furthermore, p-value analysis for the models that are shows that the only variable with coefficients[4] not statistically significant is the GDP growth rate (TX_GDP).

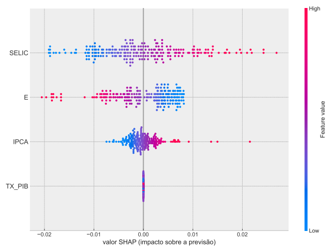
Modelo	R^2	MAPE	**coef**			
			IPCA	**TX_PIB**	**E**	**SELIC**
GLM	0.55	0.1394	0.1255*	0.0021**	-0.1464*	0.2799
Bayesian Regression	0.52	0.1441	0.0085*	0.0001**	-0.0068*	0.0238
Ridge	0.52	0.1442	0.0079*	-0.0030**	-0.0067*	0.0254
Robust	0.51	0.1501	0.0067*	-0.0118*	-0.0079**	0.0298*
Random Forest	0.94	0.0303	-	-	-	-
MLP Regressor	0.51	0.1212	-	-	-	-

Table 3: Estimated coefficients. *p-valor<0.05; **p-valor>0.05

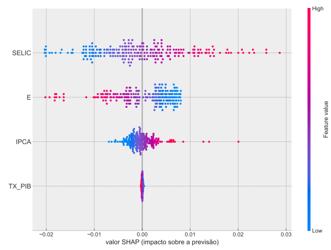
To observe the average impact of each variable for the prediction of the variable of interest, we used the SHAP method. The results obtained (Appendix Figures 2 and 3) show that all regression methods applied are consistent in terms of the order of importance of the variables, which are, respectively: (i) Selic Rate ; (ii) exchange rate; (iii) IPCA and; (iv) GDP growth rate.



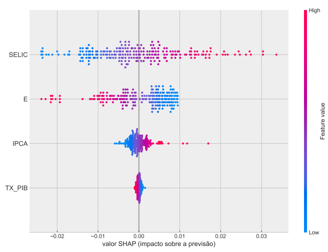
(a) MLG - N(μ, σ^2)



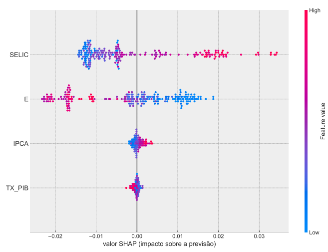
(b) Bayesian Regression



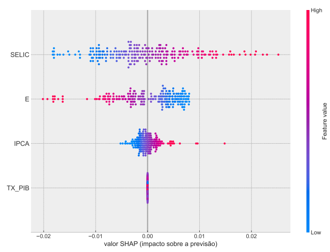
(c) *Ridge*



(d) Robust Regression



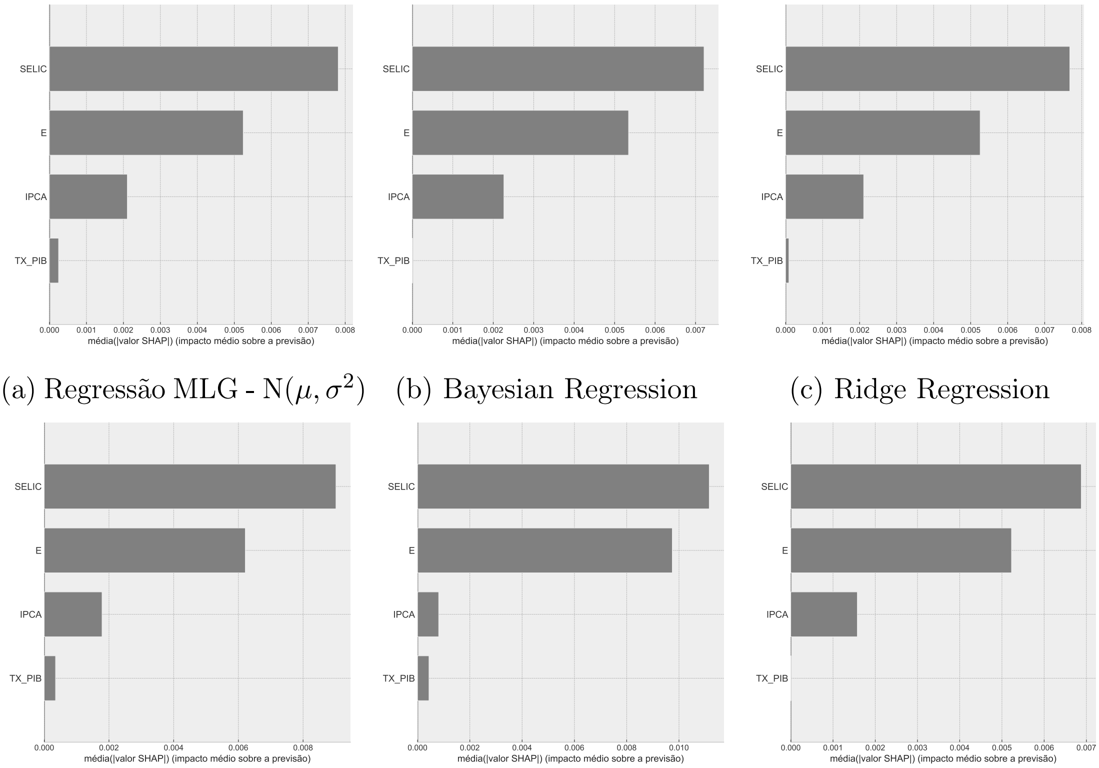
(e) *Random Forest*



(f) MLP

Figure 2: Variables impact on degree of monetization

[4]Machine learning (*Random Forest*) and artificial neural networks (MLP) methods do not have estimated coefficients. In the case of *Random Forest*, it is a combination method for a set of decision trees. And in the case of neural networks, the coefficients cannot be specified since the determination of the initial weights, an approximation of what would be the network coefficients, is performed randomly and unidentified throughout the iteration process between the nodes.



(a) Regressão MLG - N(μ, σ^2) (b) Bayesian Regression (c) Ridge Regression

(d) Robust Regression (e) Random Forest (f) MLP Regressor

Figure 3: Impacto médio das variáveis selecionadas na previsão do grau de monetização da economia brasileira

Using the analysis of variance (ANOVA) method, we verified that the mean of the SHAP values for the selected variables (SELIC, IPCA, and E) is not significant at a level of 0.05% (p-value Table 4), that is, the means are statistically equal. For the GDP growth rate, the p-value result is significant, showing that the averages found are statistically different.

	Sum of Squares	**Mean Square**	**F-value**	**p-value**
SHAP Value				
IPCA	5.657e+21	1.131e+21	1.052	0.386
PIB	1.315e+23	2.630e+22	2.776	0.0167*
E	1.140e+21	2.280e+20	0.488	0.785
SELIC	2.772e+21	2.772e+21	1.493	0.189

Table 4: Mean difference test results. *p-valor$<$0.05

4.2 Forecast Results

To measure the predictive performance of the models, simulations were performed based on the change of the arbitrary parameters used to train the model. Thus, the simulations for the MLP, LSTM and GRU models considered 5, 10 and 20 hidden layers. The model learning rate (ϵ) was considered for the simulation at 0.01 and 0.001. Furthermore, the number of epochs was fixed at 120 for each simulation and the ADAM [19] optimizer was used.

The variation in the number of nodes in the input layer seeks to assess whether a smaller error can be achieved through less deep networks. As seen in Table 5, the model with the best predictive performance (MSE 0.0071) was the RNN model with 20 nodes in the hidden layer, ϵ at 0.01. Thus, this result indicates that the use of a model with more nodes in the hidden layer can be useful for the task of predicting the degree of monetization of the Brazilian economy based on macroeconomic variables. In addition, models with a lower degree of complexity are also efficient in the forecasting exercise.

Epochs	ϵ	Hidden Nodes	MSE	Model
120	0.010	5 nodes	0.0588	RNN
120	0.010	10 nodes	**0.0071**	RNN
120	0.010	20 nodes	0.0117	RNN
120	0.001	5 nodes	0.3654	RNN
120	0.001	10 nodes	0.0223	RNN
120	0.001	20 nodes	0.0287	RNN
120	0.010	5 nodes	0.0458	LSTM
120	0.010	10 nodes	0.0302	LSTM
120	0.010	20 nodes	0.0227	LSTM
120	0.001	5 nodes	0.2240	LSTM
120	0.001	10 nodes	0.1602	LSTM
120	0.001	20 nodes	0.1100	LSTM
120	0.010	5 nodes	0.0091	GRU
120	0.010	10 nodes	0.0282	GRU
120	0.010	20 nodes	0.0216	GRU
120	0.001	5 nodes	0.0442	GRU
120	0.001	10 nodes	0.1516	GRU
120	0.001	20 nodes	0.0936	GRU

Table 5: Result of simulations

5 Conclusion

The degree of monetization of the economy can be impacted by the risk perceived by economic agents with the future behavior of the economy. Thus, this study sought to estimate the level of explainability of macroeconomic variables for the degree of monetization of the economy, taking into account that this is a variable of interest for the study of trends that justify the need to issue a digital currency by the center. Representatives of different classes of models were used as explanatory models: conventional linear, linear with regularizations, machine learning, and artificial neural networks.

By comparing neural network models to predict the degree of monetization of the economy through macroeconomic variables validated through explanatory models, the mean square error metric indicates that the *Recurrent Neural Network* model with 20 nodes in the hidden layer has the best predictive performance compared to the other architectures under analysis (*Gated Recurrent Unit* and *Long Short-Term Memory*).

The results indicate that the use of SHAP values for the selection of variables to analyze and predict the behavior of economic phenomena is effective. In this sense, future research should be developed to improve the workflow from machine learning models and their respective explanatory models, in addition to evaluating the ability to select variables for other economic and financial indicators.

References

[1] Serkan Aras and Paulo J.G. Lisboa. Explainable inflation forecasts by machine learning models. *Expert Systems with Applications*, 207:117982, 2022.

[2] Banco Central do Brasil. Sistema gerenciador de séries temporais, 2021. Online; aessado em 27 de fevereiro de 2021.

[3] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[4] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley, 2015.

[5] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[6] Marcus Buckmann, Andreas Joseph, and Helena Robertson. An interpretable machine learning workflow with an application to economic forecasting. Technical report, Bank of England, 2022.

[7] Fernando Carvalho, Luiz Ferna Paula, João Sicsú, Francisco Souza, and Rogério Studart. *Economia monetária e financeira: teoria e política*. Elsevier Brasil, 2012.

[8] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation, 2014.

[9] Kleyton Vieira Sales da Costa, Felipe Leite Coelho da Silva, and Josiane da Silva Cordeiro Coelho. Forecasting quarterly brazilian gdp: Univariate models approach, 2020.

[10] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning, 2017.

[11] Samreen Fatima and Ghulam Hussain. Statistical models of kse100 index using hybrid financial systems. *Neurocomputing*, 71(13):2742 – 2746, 2008. Artificial Neural Networks (ICANN 2006) / Engineering of Intelligent Systems (ICEIS 2006).

[12] Alex Graves. *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer-Verlag Berlin Heidelberg, 2012.

[13] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *Random Forests*, pages 587–604. Springer New York, New York, NY, 2009.

[14] John Hicks. Critical essays in monetary theory. 1967.

[15] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997.

[16] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[17] Charles C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5 – 10, 1957.

[18] Michael Y. Hu, Guoqiang (Peter) Zhang, Christine X. Jiang, and B. Eddy Patuwo. A cross-validation analysis of neural network out-of-sample performance in exchange rate forecasting. *Decision Sciences*, 30(1):197–216, 1999.

[19] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

[20] Francesca Lazzeri. *Machine Learning for Time Series Forecasting with Python*. John Wiley & Sons, 2021.

[21] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, 2017.

[22] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. Explainable ai for trees: From local explanations to global understanding, 2019.

[23] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, Dec 1943.

[24] Christoph Molnar. *Interpretable Machine Learning*. 2021.

[25] John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384, 1972.

[26] Chakradhara Panda and V. Narasimhan. Forecasting exchange rate better with artificial neural network. *Journal of Policy Modeling*, 29(2):227 – 236, 2007.

[27] Elvezio M Ronchetti and Peter J Huber. *Robust statistics.* John Wiley & Sons, 2009.

[28] D. E. Rumelhart and J. L. McClelland. *Learning Internal Representations by Error Propagation*, pages 318–362. MIT Press, 1987.

[29] L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953.

[30] Mario Henrique Simonsen and Rubens Penha Cysne. *Macroeconomia.* Atlas, 2009.

[31] Mukund Sundararajan and Amir Najmi. The many shapley values for model explanation, 2020.

[32] David West, Scott Dellana, and Jingxia Qian. Neural network ensemble strategies for financial decision applications. *Computers & Operations Research*, 32(10):2543 – 2559, 2005. Applications of Neural Networks.

[33] Mike West and Jeff Harrison. *Bayesian forecasting and dynamic models.* Springer Science & Business Media, 2006.

[34] Peter R. Winters. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342, 1960.

[35] Lean Yu, Shouyang Wang, and K.K. Lai. A novel nonlinear ensemble forecasting model incorporating glar and ann for foreign exchange rates. *Computers Operations Research*, 32(10):2523 – 2541, 2005. Applications of Neural Networks.

[36] G P Zhang and V L Berardi. Time series forecasting with neural network ensembles: an application for exchange rate prediction. *Journal of the Operational Research Society*, 52(6):652–664, 2001.

[37] Gioqinang Zhang and Michael Y. Hu. Neural network forecasting of the british pound/us dollar exchange rate. *Omega*, 26(4):495 – 506, 1998.