

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

Kleyton Vieira Sales da Costa,\* Antonio Jose Alves Junior,† Felipe Leite Coelho da Silva ‡, André de Melo Modenesi §,  
Débora Mesquita Pimentel ¶

## Resumo

Este estudo tem como objetivo avaliar a capacidade dos métodos de explicabilidade para modelos de inteligência artificial (XAI) na seleção de variáveis de resposta para os fenômenos econômicos. Para alcançar esse objetivo foi construído um fluxo de trabalho considerando modelos de aprendizado de máquina (*Random Forest* e *Multilayer Perceptron Regressor*) e modelos de regressão múltipla convencionais (Modelo Linear Generalizado, Regressão Bayesiana, Regressão Robusta e Regressão Regularizada - *Ridge*). A variável econômica dependente selecionada foi o grau de monetização da economia brasileira no período entre janeiro de 2001 e janeiro de 2021. O comportamento do grau de monetização pode ser impactado através do risco percebido pelos agentes econômicos em relação ao comportamento futuro da economia e, por esse fator, foram selecionadas variáveis usualmente utilizadas nas previsões dos agentes econômicos, sendo elas: taxa Selic, taxa de crescimento do PIB, taxa de câmbio e taxa de inflação. Os resultados encontrados através do mensuração do valor SHAP indicam que as variáveis mais importantes para prever o grau de monetização da economia são, respectivamente: a taxa Selic, a taxa de câmbio e a taxa de inflação. Por fim, aplicou-se um estudo comparativo entre modelos de redes neurais (*Recurrent Neural Network*, *Long Short-term Memory* e *Gated Recurrent Unit*) artificiais para a previsão do grau de monetização considerando as variáveis selecionadas pelo método SHAP como os valores de entrada. O estudo mostrou que o modelo *Recurrent Neural Network* tem a melhor capacidade preditiva dentre os modelos selecionados.

**Palavras-chave:** Agregados Monetários, Redes Neurais Artificiais, Interpretabilidade, Previsão

## 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.

Área: Microeconomia, Métodos Quantitativos e Finanças

**Keywords:** Monetary Aggregates, Artificial Neural Networks, Interpretability, Forecasting

**JEL Codes:** C45, C52, C53, D89

---

\*kleyton.vsc@gmail.com - Economista pela Universidade Federal Rural do Rio de Janeiro, Pesquisador na Holistic AI  
†Professor Titular do Departamento de Economia da Universidade Federal Rural do Rio de Janeiro - antonio.alves.2005@gmail.com  
‡Professor Associado do Departamento de Matemática da Universidade Federal Rural do Rio de Janeiro - felipeleiteur@ufrrj.br - Pesquisador do CNPq  
§Professor Associado do Instituto de Economia da Universidade Federal do Rio de Janeiro - amodenesi@gmail.com - Pesquisador do CNPq e da FAPERJ  
¶Professora Associada do Departamento de Economia da Universidade Federal Rural do Rio de Janeiro - pimentel.deboram@gmail.com

# 1 Introdução

O grau de monetização da economia é representado neste estudo pela razão entre o total de moeda em poder do público e depósitos à vista e o total de moeda aplicado em ativos que rendem juros. Para se calcular o grau de monetização da economia são utilizados dois componentes que pertencem aos denominados Agregados Monetários (AM): meios de pagamentos restritos e poupança financeira. Os AM são séries históricas que representam uma agregação de formas diferentes de moeda para diferentes níveis de liquidez, usualmente em uma ordenação decrescente a partir dos ativos de maior liquidez para os de menor liquidez<sup>1</sup> (Simonsen and Cysne, 2009).

A ideia de uma hierarquia estabelecida a partir do grau de liquidez dos ativos de uma economia está presente em Hicks (1967) através dos conceitos de ativos plenamente líquidos, líquidos e ilíquidos. Além da própria moeda, os ativos plenamente líquidos são aqueles que podem ser convertidos em moeda sem custos de transação significativos e sem atrasos. Os ativos líquidos são aqueles transacionados em mercados organizados e com regras de negociação bem definidas, como horários e datas agendadas para que as trocas possam ocorrer. Os agentes usualmente adquirem ativos líquidos com a expectativa de valorização futura destes, mesmo que exista algum custo de carregamento, como é o caso de obras de arte e títulos públicos. Os ativos ilíquidos são aqueles em que a sua negociação ocorre em mercados não organizados, sendo o preço de revenda menor do que o preço de aquisição, como bens de consumo duráveis em posse dos indivíduos ou equipamentos de capital em posse das empresas. Entretanto, o limiar entre o grau de liquidez dos ativos pode ser menos claro a partir da forma de organização do mercado e das consequentes características inerentes a este mercado.

As autoridades monetárias, na figura dos bancos centrais, são responsáveis pela emissão de moeda nos países que possuem uma moeda corrente própria. Entretanto, como descrito em (Carvalho et al., 2012, p.19), considera-se como emissão monetária apenas a parcela dos recursos emitidos que consta como encaixes bancários e papel-moeda em poder do público. A parcela que permanece no caixa do banco central não é legalmente considerada moeda. A composição dos meios de pagamento restritos (M1) ocorre através do somatório entre o volume de papel-moeda em poder do público e dos depósitos à vista. Estes são considerados ativos de plena liquidez, levando-se em consideração que realizam de forma completa a função de reserva de valor, podendo liquidar dívidas formadas através de contratos formais ou obrigações que a partir de transações no mercado à vista (Carvalho et al., 2012).

Com o crescimento da capacidade computacional os modelos de aprendizado de máquina se tornaram populares para aplicações em diversos campos do conhecimento, inclusive em economia. Um modelo de aprendizado de máquina é um programa que mapeia entradas e gera saídas. Por se tratarem de modelos do tipo "caixa-preta", existem barreiras para a sua utilização dada incapacidade de atender requisitos de transparência e auditoria. Em termos gerais, um algoritmo explicável ou interpretável é aquele em que as razões para uma decisão de algum modelo "caixa-preta" podem ser questionadas e explicadas de uma maneira que faça sentido para os seres humanos. Assim, a *explicabilidade* é a capacidade de explicar o que está acontecendo no algoritmo em termos de variáveis mais ou menos relevantes para a previsão, e *interpretabilidade* é a capacidade de transmitir esses resultados em termos humanamente compreensíveis (Doshi-Velez and Kim, 2017).

Neste estudo, utilizamos um modelo de explicabilidade para mensurar o impacto individual (e médio) para um conjunto de variáveis em relação a previsão de uma variável alvo. Assim, através dos resultados gerados pelos modelos é possível identificarmos quais são as variáveis mais relevantes. Levando em consideração que o grau de monetização de uma economia é uma variável de interesse para se investigar o comportamento dos agentes econômicos em relação aos meios de pagamento restritos e a poupança financeira da economia, este capítulo estima o impacto de variáveis macroeconômicas no grau de monetização da economia brasileira para o período de dezembro de 2001 a janeiro de 2021 e realiza um exercício de previsão por meio de modelos de aprendizado profundo.

---

<sup>1</sup>Nota-se que o sistema monetário e financeiro contemporâneo possui instrumentos financeiros com alta liquidez e rendimentos em juros, como é o caso das contas de poupança.

O artigo está estruturado da seguinte maneira: a Seção 2 uma breve revisão da literatura, a Seção 3 apresenta a metodologia utilizada, a Seção 4 descreve e discute os principais resultados encontrados e, por fim, a Seção 5 apresenta as conclusões e as propostas de pesquisas futuras.

## 2 Revisão da Literatura

A construção de um fluxo de trabalho proposta por [Buckmann et al. \(2022\)](#) leva em consideração a *interpretabilidade* de modelos de aprendizado de máquina como um instrumento eficiente para previsão de fenômenos econômicos. O estudo analisou os dados de empregos no Estados Unidos por meio de um conjunto de modelos de aprendizado de máquina e modelos lineares convencionais, concluindo que os primeiros tiveram um melhor desempenho. Além disso, valores *Shapley* foram aplicados para identificar comportamentos não lineares que tenham significância econômica.

O estudo proposto por [Aras and Lisboa \(2022\)](#) utiliza valores SHAP com a finalidade de selecionar variáveis que tenham a capacidade de prever a inflação em uma economia emergente (Turquia) de forma mais acurada. O estudo propôs a comparação entre modelos de fatores e modelos de aprendizado de máquina, obtendo como resultado um melhor desempenho dos modelos de aprendizado de máquina.

A previsão de séries temporais possui abordagens metodológicas que se apresentam através de diferentes classes de modelos e métodos. Uma abordagem univariada pode ser construída através dos modelos clássicos, amplamente aplicados na previsão de séries econômicas. Os mais conhecidos dessa classe são os modelos de Box & Jenkins ([Box et al., 2015](#)), que em uma forma ampliada leva em consideração a sazonalidade que pode ser identificada em uma série através da função de autocorrelação. Os modelos de suavização exponencial [[Holt \(1957\)](#), [Winters \(1960\)](#)] também são abordagens metodológicas com aplicações diversas aos dados reais, tomando equações de suavização de nível, tendência e sazonalidade e que são capazes de se ajustar de forma significativa a dados trimestrais. A classe de modelos de estado de espaços [West and Harrison \(2006\)](#) é outra que possui grande relevância e aplicabilidade para dados econômicos e financeiros. Um expoente dessa classe são os modelos lineares dinâmicos que, por meio de um arcabouço Bayesiano robusto e sofisticado, são capazes de gerar resultados expressivos, como proposto em [da Costa et al. \(2020\)](#). Além disso, os modelos de redes neurais artificiais ganham cada vez mais espaço para a análise e previsão de séries temporais.

As aplicações de redes neurais para a previsão de taxas de câmbio possuem resultados que são influenciados por problemas ligados a seleção dos dados e a variação amostral. Grande parte das aplicações adotam uma prática de validação do modelo que consiste em separar os dados entre um conjunto de treino (*in-sample*) e um conjunto de teste (*out-sample*). [Hu et al. \(1999\)](#) observa dois problemas principais nessa maneira de selecionar os dados para treino e teste. É introduzido viés no modelo quando as características observadas do conjunto de treino são diferentes das observadas no conjunto de teste. E não é levado em consideração os efeitos que o tamanho do conjunto de treino exerce sobre o modelo. Dessa maneira, a proposta do autor é utilizar uma metodologia de *cross-validation* com a finalidade de analisar o desempenho *out-sample* das redes neurais. O estudo utiliza redes neurais aplicadas com a finalidade de prever os dados semanais da taxa de câmbio entre as divisas GBP e USD. Os resultados obtidos apontam que a acurácia de redes neurais não possui uma sensibilidade significativa em relação a variação amostral. Outro resultado encontrado é o de que o modelo de redes neurais realiza previsões com melhor desempenho para o curto prazo, obtendo também resultados mais significativos do que os observados no *random walk*.

O estudo proposto por [Zhang and Hu \(1998\)](#) examina como o número de camadas ocultas, camadas de entrada e o tamanho da amostra impactam o desempenho de previsões dentro da amostra e fora da amostra. A taxa de câmbio analisada é entre GBP/USD, sendo os dados diários compreendidos para o período entre 1976 a 1993. O modelo utilizado foi o comumente utilizado *multilayer perceptron*. Como forma de estabilizar a série, os autores optaram por utilizar uma transformação logarítmica. O estudo aponta que redes neurais possuem resultados mais significativos do que modelos lineares, principalmente no curto prazo. O número de entradas possui tanto impacto no desempenho do modelo quando o número de camadas ocultas, assim

como um maior número de observações contribui para a redução dos erros de previsão. As previsões foram geradas semanalmente, apontando que a seleção das entradas da rede neural e da arquitetura são fatores significativos para a eficácia preditiva.

Um aplicação para a previsão para a taxa de câmbio foi proposta por [Zhang and Berardi \(2001\)](#). Autores utilizam uma abordagem que leva em consideração a "união de modelos de redes neurais" (*neural networks ensembles*) com a finalidade de aprimorar o desempenho da rede neural, sendo uma abordagem distinta da ideia padrão de se escolher um único melhor modelo. Assim, o artigo apresenta a união das redes neurais para a previsão da taxa de câmbio. Os resultados apontam que a combinação das redes neurais gerou uma melhoria significativa nos resultados em relação as redes neurais individuais. Mas por outro lado, os resultados não obtiveram resultados significativamente melhores do que o modelo de *random walk*.

A metodologia proposta em [Yu et al. \(2005\)](#) propõe uma combinação não-linear entre um modelo linear autoregressivo generalizado (GLAR) e um modelo de redes neurais artificiais (ANN), buscando aumentar a acurácia preditiva. O modelo híbrido foi comparado com as suas versões isoladas e aplicados aos dados das taxas de câmbio DEM/USD, GBP/USD e JPY/USD em valores mensais, compreendidos para o período entre janeiro de 1971 a dezembro de 2000. Os resultados apontam que o modelo híbrido proposto (GLAR-ANN) obteve um resultado de previsão superior aos modelos GLAR e ANN.

A literatura também é composta por estudos que utilizam modelos de redes neurais artificiais para previsão de ativos financeiros. [West et al. \(2005\)](#) investiga a utilização de uma combinação entre modelos de redes neurais com a finalidade de melhorar a capacidade preditiva dos modelos. Para a análise empírica foram utilizados os dados de crédito da Austrália e da Alemanha, além de dados de falência fornecidos pela agência de classificação *Standard and Poor's*. Os resultados encontrados apontam que a combinação dos modelos obteve um melhor desempenho do que a estratégia de um único melhor modelo. Outra aplicação de modelos híbridos/combinações foi elaborada em [Fatima and Hussain \(2008\)](#). Neste trabalho, a metodologia utilizada consiste em um modelo híbrido de sistemas financeiros (HFSs) combinando redes neurais artificiais com modelos ARIMA ou ARCH/GARCH. Para a análise foram utilizados dados do índice da bolsa de valores Karachi (KSE100), em valores diários, para o período compreendido entre 01 de janeiro de 2000 e 18 de outubro de 2002. O modelo híbrido ( $ANN_{ARCH/GARCH}$ ) apresentou o melhor desempenho em comparação aos outros modelos levados em consideração no estudo (ANN, ARIMA e GARCH).

### 3 Metodologia

Esta seção apresenta a metodologia utilizada neste trabalho. Foram implementadas duas abordagens de modelagem: modelos explicativos e modelos preditivos para o grau de monetização da economia brasileira. Os modelos explicativos tem a finalidade de determinar quais variáveis são significativas para explicar o grau de monetização. Além disso, um modelo de interpretabilidade foi aplicado para identificar quais variáveis possuem maior impacto para os resultados dos modelos. Nesse estágio de identificação foram considerados modelos de aprendizado de máquina, aprendizado profundo e modelos lineares generalizados que estão descritos na Subseção 3.2.

Após a validação das variáveis selecionadas, foram implementados modelos de redes neurais artificiais para previsão de séries temporais. Nesse estágio se realizou um exercício de simulação a partir de diferentes configurações para os parâmetros iniciais. O desempenho dos modelos foi comparado através de uma abordagem convencional através de métricas de precisão.

#### 3.1 Descrição dos Dados

Para a condução dos experimentos foram selecionadas as variáveis listadas abaixo, descritas na Tabela 4 e apresentadas na Figura 1. Todos os dados foram obtidos através do Sistema Gerenciador de Séries Temporais do BCB ([Banco Central do Brasil, 2021](#)). Observa-se (Tabela 4) que as variáveis possuem 229 observações, compreendendo o período entre janeiro de 2001 e janeiro de 2021. As variáveis selecionadas são: M1 - Meios de pagamento restritos (saldo em final de período) - u.m.c. (mil); M4 - Poupança financeira. Meios

de pagamento amplos - (saldo em final de período) - u.m.c. (mil); IPCA - Índice de preços ao consumidor amplo; SELIC - Taxa de juros, Selic acumulada no mês; PIB - Taxa de crescimento do produto interno bruto mensal a valores correntes; E - Taxa de câmbio para o real em relação ao dólar americano.

### 3.2 Modelos Explicativos para o Grau de Monetização

Para investigar os elementos que afetam o grau de monetização da economia brasileira e o sinal desses efeitos, utilizou-se uma estratégia de regressão com a finalidade de mensurar o nível de explicabilidade que variáveis macroeconômicas exercem sobre o fenômeno analisado. Os modelos explicativos propostos foram construídos a partir de seis abordagens: (i) modelo linear generalizado (Gaussiano); (ii) modelo linear generalizado (Gamma); (iii) regressão bayesiana; (iv) regressão *Ridge*; (v) regressão robusta e; (iii) *Multilayer Perceptron Regressor*.

Para mensurar o impacto (e o impacto médio) das variáveis selecionadas na capacidade de explicação dos modelos propostos foi utilizado o modelo SHAP (Lundberg et al., 2019), uma estratégia computacional baseada no *valor de Shapley*. A utilização de valores de Shapley para identificar quais variáveis são mais relevantes para explicar um fenômeno econômico a partir de métodos de aprendizado de máquina é uma alternativa aos métodos econométricos tradicionais (Buckmann et al., 2022).

**Modelo Linear Generalizado** Foram considerados para aplicação a classe de modelos lineares generalizados, uma extensão dos modelos lineares simples proposta por Nelder and Wedderburn (1972). O modelo considera uma combinação a partir de um componente sistemático e um componente aleatório. Assim, a extensão a partir de um modelo linear convencional ocorre através de um função de ligação ( $\theta$ ) aplicada através de uma combinação linear entre os preditores ( $\hat{y} = \sum_{i=1}^m \beta_i x$ ) e o conjunto de observações  $x_i = (1, \dots, i)$ . Esse processo pode ser expresso como

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

em que  $\beta_i$  são os coeficientes associados às variáveis, com  $\beta_i = (1, 2, \dots, i)$ . A função de ligação  $\theta = (x_i \beta_i)$  conecta o parâmetro  $\theta$  da distribuição  $z$  com os valores observados na regressão linear. Para o estudo foram considerada uma distribuição Gaussiana -  $\sum(z - \mu^2)/\sigma^2$ . Utilizando a distribuição Gaussiana obtemos um modelo de regressão linear com erros que seguem  $N(\theta, \sigma^2)$ .

**Regressão Regularizada por Ridge** Os modelo de regressão que utilizam regularização são estratégias para lidar, em essência, com os problemas de sobre-ajuste, ou seja, quando os modelos são capazes de se ajustar aos dados observados de forma expressiva mas não realizam boas previsões, considerando determinado conjunto de treino  $x_i = (1, \dots, n)$  e teste  $y_i = (n + 1, \dots, m)$ . Neste estudo foi aplicada a abordagem de regressão com regularização *Ridge* ( $L_2$ ). Uma penalização do tipo  $L_2$  adota uma restrição nos coeficientes através de um fator de penalização. A estratégia de regularização considerada no método  $L_2$  utiliza os quadrados dos coeficientes (Hoerl and Kennard, 1970) estimados pelo modelo de regressão, ou seja,

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

em que  $\lambda$  determina o tamanho da restrição que será imposta na regressão. Quanto maior for o valor de  $\lambda$  maior será a penalização da função de otimização, considerando que  $\lambda = 0$  representa um modelo linear simples. Além de reduzir a complexidade do modelo (e o sobre-ajuste), a regularização  $L_1$  reduz a multicolinearidade.

**Regressão Bayesiana** A regressão bayesiana é utilizada para adicionar parâmetros de regularização no processo de estimação através de um processo de iteração. Essa abordagem bayesiana para uma regressão linear contorna o problema de sobre-ajuste observado através da estimação por máxima verossimilhança

(Bishop and Nasrabadi, 2006). Uma abordagem bayesiana para a regressão *Ridge* é construída a partir da minimização de  $\hat{\beta}$  como sendo a média *posteriori* de um modelo em que

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

**Regressão Robusta** A regressão robusta (Ronchetti and Huber, 2009) tem como objetivo ajustar uma regressão em dados que tenham um comportamento não padronizado, ou seja, tenham presença de *ouliers* ou erros de especificação. Neste estudo foi utilizada uma estratégia de estimação do tipo Huber ( $H_\epsilon$ ), aplicando-se uma função de erro linear para os dados da amostra que são identificados como *ouliers*. A determinação se uma amostra não é uma observação incomum ocorre se o erro absoluto da amostra é menor do que um *threshold*. Dessa maneira, a função de erro para um estimador Huber dada por,

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

considerando a função do estimador como,

$$H_\epsilon(z) = \begin{cases} z^2, & \text{se } |z| < \epsilon \\ 2\epsilon|z| - \epsilon^2, & \text{caso contrário.} \end{cases} \quad (5)$$

**Random Forest** Uma *Random Forest* se caracteriza como um algoritmo para a combinação de um conjunto de *decision trees* em que cada árvore depende dos valores de um vetor aleatório com amostras independentes e com a mesma distribuição (Breiman, 2001). A modificação decisiva em relação ao método originário (*Bagging*) está no desenvolvimento de um número expressivo de *decision trees* que não são correlacionadas e calcular a média delas a partir desse ponto. O Algoritmo 1 apresenta os passos que devem ser seguidos para que uma *Random Forest* seja aplicada com a finalidade de regressão.

---

**Algorithm 1** Algoritmo para a construção de uma *Random Forest* (Hastie et al., 2009)

---

1. considerando  $b = (1, \dots, B)$ 
  - (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  $T_b^B$

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** Um modelo *Multilayer Perceptron Regressor* (MLPR) realiza o processo de treinamento através de *backpropagation* sem que seja aplicada uma função de ativação na camada de saída. Na formulação utilizada neste estudo uma função de perda do tipo erro quadrático médio foi considerada. Além disso, as saídas são definidas a partir de um conjunto de valores contínuos. A definição do modelo se dá da seguinte forma: dado um conjunto de treino  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  em que  $x_i \in \mathbb{R}^n$  e  $y_i \in \{0, 1\}$ , um modelo MLPR com uma camada oculta realiza o processo de aprendizado através da função

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

em que  $W_1 \in \mathbb{R}^m$  e os parâmetros do modelo são  $W_2, b_1, b_2 \in \mathbb{R}$ .  $W_1$  e  $W_2$  são, respectivamente, os pesos da camada de entrada e da camada oculta.  $b_1$  e  $b_2$  são, respectivamente os vieses da camada oculta e da

camada de saída. A função  $g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$  é uma função de ativação, neste estudo, do tipo tan hiperbólica que é dada por,

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

Por fim, foi utilizado como função de perda ( $\mathcal{L}$ ) o erro quadrático médio. Assim, a função foi definida como

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

em que a partir de pesos iniciais definidos aleatoriamente, o modelo minimiza  $\mathcal{L}$  a partir de atualizações repetidas para os pesos. Com o cálculo de  $\mathcal{L}$ , cada parâmetro de peso é atualizado com o intuito de diminuir a perda.

**SHAP - SHapley Additive exPlanations** A definição do impacto das variáveis de interesse utilizadas como entrada para um modelo convencional ou de aprendizado de máquina, pode ser alcançada através de métodos de explicabilidade. A finalidade do método SHAP é explicar a previsão de uma variável dependente a partir da contribuição de cada variável independente selecionada. A inovação que método apresenta em relação ao valor Shapley ( $\phi_i(v)$ )<sup>2</sup> é a representação dada para o valor a partir de um modelo linear descrito por meio de um método aditivo para as variáveis. O método pode ser descrito como,

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

em que  $\gamma$  é o modelo de explicação,  $z' \in \{0, 1\}^C$  é o vetor de coalizão,  $C$  é o tamanho máximo da coalizão e  $\phi_i(v) \in \mathbb{R}$  é valor Shapley.

Tomando como entrada uma função de conjunto  $v : 2^n \rightarrow \mathbb{R}$ , podemos definir  $\phi_i(v)$  para uma variável específica  $i$  [(Sundararajan and Najmi, 2020), (Molnar, 2021), (Lundberg and Lee, 2017)] como sua contribuição para o pagamento através da média ponderada de todas as combinações possíveis

$$\phi_i(v) = \sum_{F \subseteq \{x_1, \dots, x_n\} \setminus \{x_i\}} \frac{|F|! (n - |F| - 1)!}{n!} (v(F \cup \{x_i\}) - v(F)) \quad (10)$$

em que  $n$  é o número de variáveis,  $F$  é o subconjunto das variáveis utilizadas no modelo, e  $x$  é o vetor de valores para as variáveis da amostra a ser explicada.

### 3.3 Modelos Preditivos para o Grau de Monetização

Para analisar a capacidade de previsão dos modelos de redes neurais artificiais para o grau de monetização da economia brasileira foram utilizadas os modelos *Recurrent Neural Network* (RNN), *Long Short-Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU). Os modelos de redes neurais artificias surgem como uma tentativa de aproximar o funcionamento de um cérebro biológico através de uma função matemática<sup>3</sup>. Nesse tipo de modelagem usualmente são consideradas três camadas de operação: a camada de entrada, a camada oculta e a camada de saída.

<sup>2</sup>O valor de Shapley (Shapley, 1953) tornou-se a principal contribuição para a teoria dos jogos cooperativos por alocar pagamentos a jogadores dependendo de sua contribuição para o total. Quando consideramos um contexto de Aprendizado de Máquina, cada variável é um jogador que participa do jogo; a previsão é a recompensa e o valor de Shapley comunica como a contribuição de cada variável (pagamentos) pode ser difundida entre as variáveis.

<sup>3</sup>Ver a discussão seminal de Rumelhart and McClelland (1987) e McCulloch and Pitts (1943).

**Recurrent Neural Network** As *recurrent neural networks* (RNNs) vêm ganhando mais atenção através do aumento do poder computacional das unidades de processamento gráfico (GPUs) (Lazzeri, 2021). São úteis para dados de séries temporais considerando que cada neurônio pode usar sua "memória" interna para manter as informações da entrada anterior. O problema fundamental em uma RNN simples é a incapacidade de capturar as dependências de longo prazo em uma série, ou seja, as conexões recorrentes permitem que uma "memória" de entradas anteriores persista no estado interno da rede, influenciando a saída da rede (Graves, 2012). Podemos considerar que a *forward propagation* em uma RNN é semelhante ao observado em uma MLP com uma única camada oculta, entretanto as funções de ativação atingem a camada oculta a partir da entrada externa atual e a ativação da camada oculta do intervalo de tempo anterior. A notação do modelo considerando  $i$  o número de unidades de entrada,  $H$  entradas ocultas e  $K$  unidades de saída;  $x_i^t$  o valor da entrada  $i$  no momento  $t$ ; a entrada da rede  $a_h^t$  e a ativação final  $b_h^t$ , ambas para a unidade  $h$  no momento  $t$ . As unidades ocultas são dadas por,

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

**Long Short-term Memory** O modelo LSTM é um tipo especial de *Recurrent Neural Network* (RNN), desenvolvido por Hochreiter and Schmidhuber (1997). Como observado anteriormente, as RNNs possuem uma arquitetura útil para dados organizados em séries temporais, uma vez que cada nó utiliza a sua "memória" interna para preservar a informação que pertence a entrada anterior. Entretanto, um dos problemas de maior relevância para uma RNN simples está no *problema de desaparecimento do gradiente*, sendo esta uma incapacidade de captura de dependência de longo prazo observadas na série temporal, isto é, as conexões recorrentes permitem que uma "memória" de entradas anteriores persista no estado interno da rede, influenciando a saída da rede Graves (2012). Assim, a LSTM é uma estratégia de solução para este problema, uma vez que ela considera as dependências de longo prazo que estão presentes na série temporal.

Como descrito em ?, para cada elemento na sequência de entrada, cada camada irá computar as equações

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

em que  $h_t$  o estado oculto para o tempo  $t$ ,  $c_t$  é a célula de estado para o tempo  $t$ ,  $x_t$  é a entrada para o tempo  $t$ ,  $h_{t-1}$  é o estado oculto da camada no tempo  $t - 1$  ou o estado inicial para o tempo 0, e  $i_t$ ,  $f_t$ ,  $g_t$ ,  $o_t$  são os *gates* de entrada, esquecimento, célula e saída.  $\sigma$  é a função de ativação (sigmoide), e  $\odot$  é o produto de Hadamard. Considerando um LSTM com múltiplas camadas, a entrada  $x_t^{(l)}$  para a  $l$ -ésima camada ( $l \geq 2$ ) é o estado oculto  $h_t^{(l-1)}$  da camada anterior multiplicada pelo *dropout*  $\delta_t^{(l-1)}$ , em que cada  $\delta_t^{(l-1)}$  é uma variável aleatória de Bernoulli em que 0 é a probabilidade do *dropout*.

**Gated Recurrent Unit** Proposto por Cho et al. (2014), o modelo GRU possui apenas dois *gates*: um *gate* de redefinição e um *gate* de atualização. Assim como um modelo LSTM, um GRU busca resolver o *problema de desaparecimento do gradiente*. Isso ocorre através de dois vetores (*gates*) que decidem quais informações irão ser levadas até a saída e quais não serão. Dessa maneira, a estrutura do modelo permite capturar a dependência de um volume considerável de dados sem que as informações anteriores do modelo sejam descartadas, e isto de forma adaptativa. Para cada elemento da sequência de entrada, cada camada irá computar



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

em que  $h_t$  é a camada de estado no tempo  $t$ ,  $x_t$  é a entrada no tempo  $t$ ,  $h_{(t-1)}$  é o estado oculto da camada no tempo  $t - 1$  ou o estado oculto inicial no tempo 0, e  $r_t$ ,  $z_t$ ,  $n_t$  são os *gates* de reinício, atualização e novos *gates*.  $\sigma$  é a função de ativação (sigmoide), e  $\odot$  é o produto de Hadamard. Em um GRU de multicamadas, a entrada  $x_t^{(l)}$  da  $l$ -ésima camada ( $l \geq 2$ ) é o estado oculto  $h_t^{(l-1)}$  da camada anterior multiplicado pelo *dropout*  $\delta_t^{(l-1)}$  em que  $\delta_t^{(l-1)}$  é uma variável aleatória Bernoulli em que 0 é a probabilidade do *dropout*.

## 4 Resultados e Discussão

Esta seção apresenta a discussão e os principais resultados encontrados neste estudo. O primeiro tema a ser analisado diz respeito a regressão múltipla estimada considerando o grau de monetização da economia como variável dependente e tendo como variáveis explicativas a taxa de crescimento produto interno produto mensal para o Brasil, a taxa de inflação mensal capturada através do índice de preços ao consumidor amplo (IPCA), a taxa de câmbio para o real brasileiro em relação ao dólar americano e a taxa básica de juros (Selic). Após a análise da regressão, apresenta-se o desempenho preditivo para os modelos de aprendizado profundo considerados.

### 4.1 Estimação do Grau de Monetização

A composição da poupança financeira (M4) de uma economia é formada pela soma dos títulos públicos de alta liquidez e dos meios de pagamento ampliados. Por sua vez, os meios de pagamentos ampliados são determinados através da soma entre os meios de pagamento restritos e os depósitos especiais remunerados, os depósitos de poupança, os títulos emitidos por instituições depositárias, as quotas dos fundos de renda fixa e as operações compromissadas registradas no Selic.

O grau de monetização da economia pode ser influenciado através de um movimento de busca por segurança por parte dos agentes econômicos, isto é, quando há aumento do risco relacionado ao comportamento futuro da economia os agentes econômicos buscam proteger o seu capital através de ativos mais ou menos líquidos, a depender da característica do fenômeno que impactará o comportamento econômico esperado. Dessa maneira, a Tabela 1 sintetiza o efeito esperado de certas variáveis em relação a M1 e M4 e, conseqüentemente, no grau de monetização da economia.

### Comportamento Esperado

Quando há um *aumento da inflação* se supõe que os agentes tendem a migrar de ativos mais líquidos (que não rendem juros) para ativos menos líquidos (que rendem juros). Assim, ocorre uma redução nos meios de pagamento restritos e uma expansão da poupança financeira, provocando o efeito esperado de uma redução no grau de monetização da economia.

Quando há um *aumento da taxa de câmbio* se supõe que os agentes tendem a migrar de ativos mais líquidos (que não rendem juros) para ativos menos líquidos (que rendem juros). Assim, ocorre uma redução nos meios de pagamento restritos e uma expansão da poupança financeira, provocando o efeito esperado de uma redução no grau de monetização da economia.

Quando há um *aumento da taxa de crescimento do produto interno bruto* se supõe que os agentes tendem a migrar de ativos menos líquidos (que rendem juros) para ativos mais líquidos (que não rendem juros). Assim, ocorre uma expansão nos meios de pagamento restritos e uma redução da poupança financeira, provocando o efeito esperado de uma redução no grau de monetização da economia.

Quando há um *aumento da taxa Selic* se supõe que os agentes tendem a migrar de ativos mais líquidos (que não rendem juros) para ativos menos líquidos (que rendem juros). Assim, ocorre uma redução nos meios de pagamento restritos e uma expansão da poupança financeira, provocando o efeito esperado de uma redução no grau de monetização da economia.

Tabela 1: Comportamento esperado do grau de monetização a partir da dinâmica de variáveis macroeconômicas.

Para mensurar o grau de monetização da economia brasileira foram utilizadas quatro variáveis independentes estimadas através do método de mínimos quadrados ordinários. O coeficiente de determinação múltipla ( $R^2$ ) apresenta resultados consistentes entre as diferentes estratégias de regressão adotadas. A Tabela 2 mostra que, exceto para o método *Random Forest*, o resultado observado para o coeficiente de determinação varia entre 0.51 e 0.55, ou seja, um resultado razoável considerando a utilização de um número parcimonioso de variáveis explicativas. Além disso, através do erro absoluto percentual médio (MAPE) aplicado para o resultado de ajuste do modelo mostra que os resultados possuem uma diferença pequena em valores absolutos, variando entre 0.1212 e 0.1501. As duas métricas mostram que o método *Random Forest* possui um melhor desempenho, 0.94 para o  $R^2$  e 0.0303 para o MAPE. Além disso, análise do p-valor para os modelos que são mostra que a única variável com coeficientes<sup>4</sup> não significativos estatisticamente é a taxa de crescimento do PIB (TX\_PIB).

Modelo	$R^2$	MAPE	coef			
			IPCA	TX_PIB	E	SELIC
<b>Modelo Linear Generalizado</b>	0.55	0.1394	0.1255*	0.0021**	-0.1464*	0.2799
<b>Regressão Bayesiana</b>	0.52	0.1441	0.0085*	0.0001**	-0.0068*	0.0238
<b>Regressão Ridge</b>	0.52	0.1442	0.0079*	-0.0030**	-0.0067*	0.0254
<b>Regressão Robusta</b>	0.51	0.1501	0.0067*	-0.0118*	-0.0079**	0.0298*
<i>Random Forest</i>	0.94	0.0303	-	-	-	-
<i>Multilayer Perceptron</i>	0.51	0.1212	-	-	-	-

Tabela 2: Coeficientes estimados para os modelos e métricas de precisão. \*p-valor<0.05; \*\*p-valor>0.05

<sup>4</sup>Os métodos de aprendizado de máquina (*Random Forest*) e de redes neurais artificiais (MLP) não possuem coeficientes estimados. No caso do *Random Forest*, trata-se de um método de combinação para um conjunto de árvores de decisão. E no caso das redes neurais, os coeficientes não podem ser especificados uma vez que a determinação dos pesos iniciais, uma aproximação do que seriam os coeficientes da rede, é realizada de forma aleatória e não identificado ao longo do processo de iteração entre os nós.

Para observar o impacto médio de cada variável para a previsão da variável de interesse utilizamos o método SHAP. Os resultados obtidos (Apêndice Figuras 2 e 3) mostram que todos os métodos de regressão aplicados são consistentes quanto a ordem de importância das variáveis, sendo elas, respectivamente: (i) Taxa Selic; (ii) taxa de câmbio; (iii) IPCA e; (iv) taxa de crescimento do PIB.

A partir do método de análise de variância (ANOVA) verificamos que a média dos valores SHAP para as variáveis selecionadas (SELIC, IPCA e E) não é significativa a um nível de 0.05% (p-valor Tabela 3), ou seja, as médias são estatisticamente iguais. Para a taxa de crescimento do PIB o resultado do p-valor é significativo, mostrando que as médias encontradas são estatisticamente diferentes.

	Soma dos Quadrados	Quadrado Médio	Estatística F	p-valor
<i>Valor SHAP</i>				
<b>IPCA</b>	5.657e+21	1.131e+21	1.052	0.386**
<b>PIB</b>	1.315e+23	2.630e+22	2.776	0.0167*
<b>E</b>	1.140e+21	2.280e+20	0.488	0.785**
<b>SELIC</b>	2.772e+21	2.772e+21	1.493	0.189**

Tabela 3: Teste de diferenças entre médias para os resultados de ajuste e para a média do valor SHAP. \*p-valor<0.05; \*\*p-valor>0.05

## 4.2 Desempenho Preditivo

Para mensurar o desempenho preditivo dos modelos foram realizadas simulações a partir da mudança dos parâmetros arbitrários utilizados para o treinamento do modelo. Assim, as simulações para os modelos MLP, LSTM e GRU consideraram 5, 10 e 20 camadas ocultas. A taxa de aprendizado do modelo ( $\epsilon$ ) foi considerada para a simulação em 0.01 e 0.001. Além disso, o número de épocas foi fixado em 120 para cada simulação e otimizador ADAM (Kingma and Ba, 2014) foi utilizado.

A variação no número de nós na camada de entrada busca avaliar se um erro menor pode ser alcançado através de redes menos profundas. Como observa-se na Tabela 5, o modelo com o melhor desempenho preditivo (MSE 0.0071) foi o modelo RNN com 20 nós na camada oculta,  $\epsilon$  em 0.01. Dessa maneira, este resultado indica que a utilização de um modelo com mais nós na camada oculta pode ser útil para a tarefa de previsão do grau de monetização da economia brasileira a partir de variáveis macroeconômicas. Além disso, modelos com um menor grau de complexidade também se mostram eficientes no exercício de previsão.

## 5 Conclusão

O grau de monetização da economia pode ser impactado através do risco percebido pelos agentes econômicos em relação ao comportamento futuro da economia. Assim, este estudo buscou estimar o nível de explicabilidade de variáveis macroeconômicas para o grau de monetização da economia, levando-se em consideração que esta é uma variável de interesse para o estudo das tendências que justificam a necessidade de emissão de uma moeda digital por parte do central. Foram utilizados para como modelos explicativos representantes de diferentes classes de modelos: lineares convencionais, lineares com regularizações, aprendizado de máquina e redes neurais artificiais.

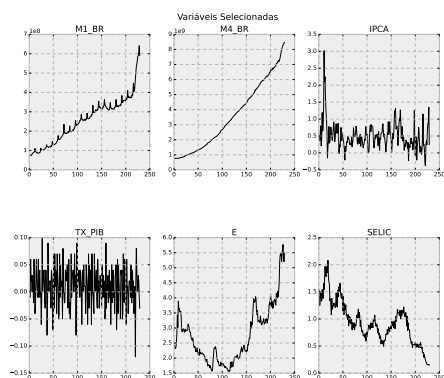
A análise mostrou que as variáveis selecionadas (IPCA, taxa de crescimento do PIB, SELIC e taxa de câmbio) possuem uma capacidade razoável para explicar o grau de monetização da economia, com um coeficiente de determinação múltipla variando entre 0.51 e 0.55, desconsiderando o resultado fora da curva do método de *Random Forest*, 0.94. A partir das hipóteses construídas para o comportamento do grau de monetização da economia, os modelos avaliados apresentaram como resultado que as variáveis SELIC, IPCA e taxa de câmbio são estatisticamente significativas para explicar a variação esperada no grau de monetização

da economia brasileira, além de possuírem mais importância para os valores ajustados pelo modelo, como observado por meio do valor SHAP.

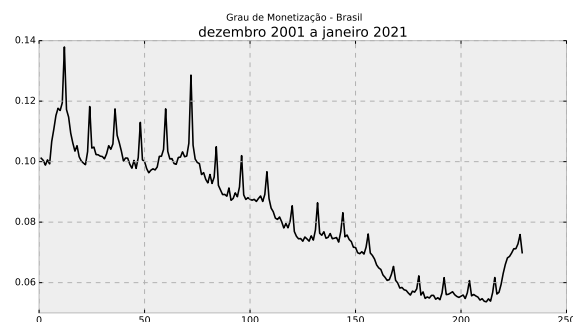
Através da comparação entre modelos de redes neurais com a finalidade de realizar a previsão do grau de monetização da economia por meio de variáveis macroeconômicas validadas através dos modelos explicativos, a métrica de erro quadrático médio indica que o modelo *Recurrent Neural Network* com 20 nós na camada oculta possui o melhor desempenho preditivo em relação as outras arquiteturas em análise (*Gated Recurrent Unit* e *Long Short-Term Memory*).

Os resultados apontam que a utilização de valores SHAP para seleção de variáveis com o intuito de analisar e prever o comportamento de fenômenos econômicos é eficaz. Nesse sentido, pesquisas futuras devem ser elaborados com o objetivo de aprimorar o fluxo de trabalho a partir de modelos de aprendizado de máquina e seus respectivos modelos de explicabilidade, além de avaliar a capacidade de selecionar variáveis para outros indicadores econômicos e financeiros.

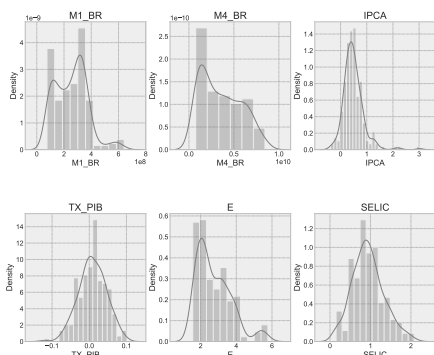
# A Variáveis Seleccionadas



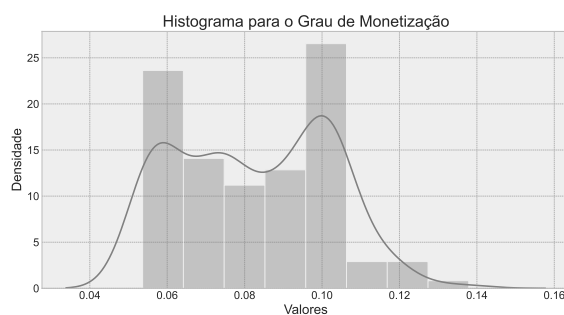
(a) Variáveis seleccionadas



(b) Grau de monetização



(c) Histograma variáveis seleccionadas



(d) Histograma variáveis seleccionadas

Figura 1: Variáveis seleccionadas e grau de monetização da economia brasileira entre dezembro de 2001 e janeiro de 2021

	M1	M4	IPCA	PIB	E	SELIC
# Observações	229	229	229	229	229	229
Média	254412701.2620	3560907502.4934	0.4925	0.0083	2.7555	0.9401
Desvio Padrão	118050957.9182	2221602631.0116	0.3897	0.0370	0.9576	0.3858
Mínimo	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
Máximo	641107547.0000	8454294566.0000	3.0200	0.1015	5.7700	2.0800

Tabela 4: Resumo estatístico para as variáveis em análise

## B Valores SHAP para os Modelos Explicativos

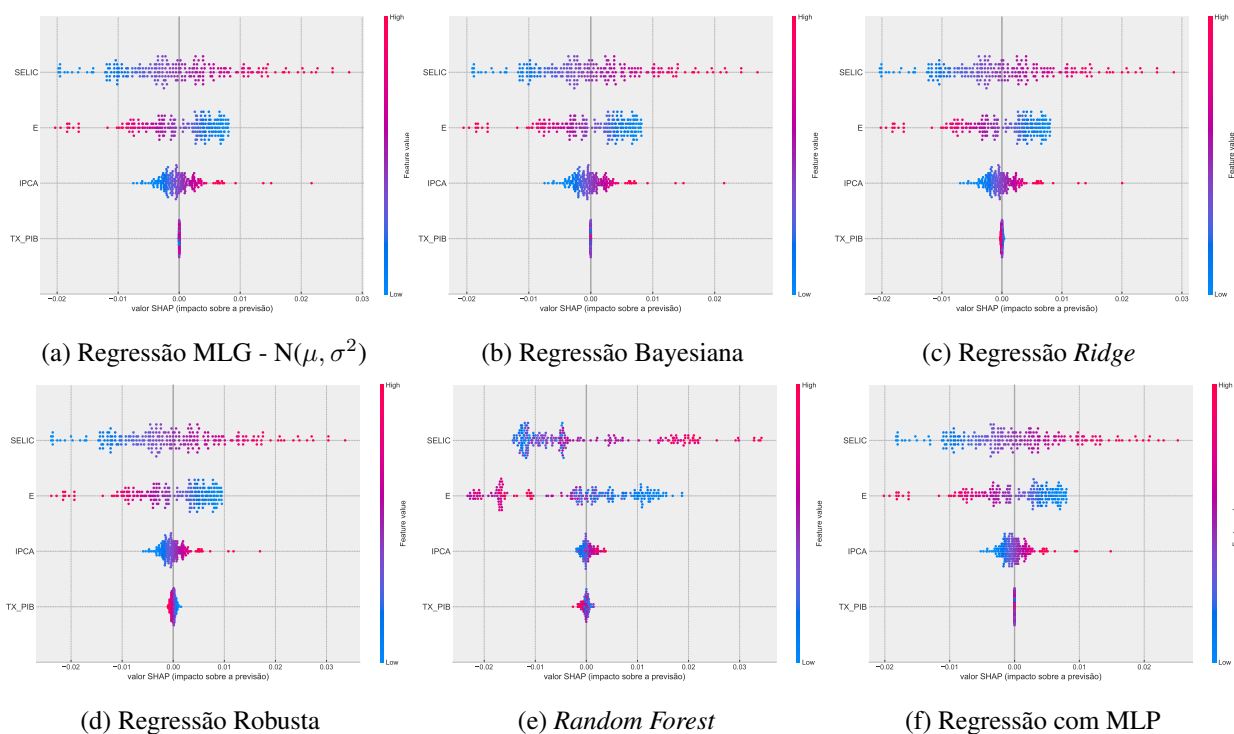


Figura 2: Impacto das variáveis selecionadas na previsão do grau de monetização da economia brasileira

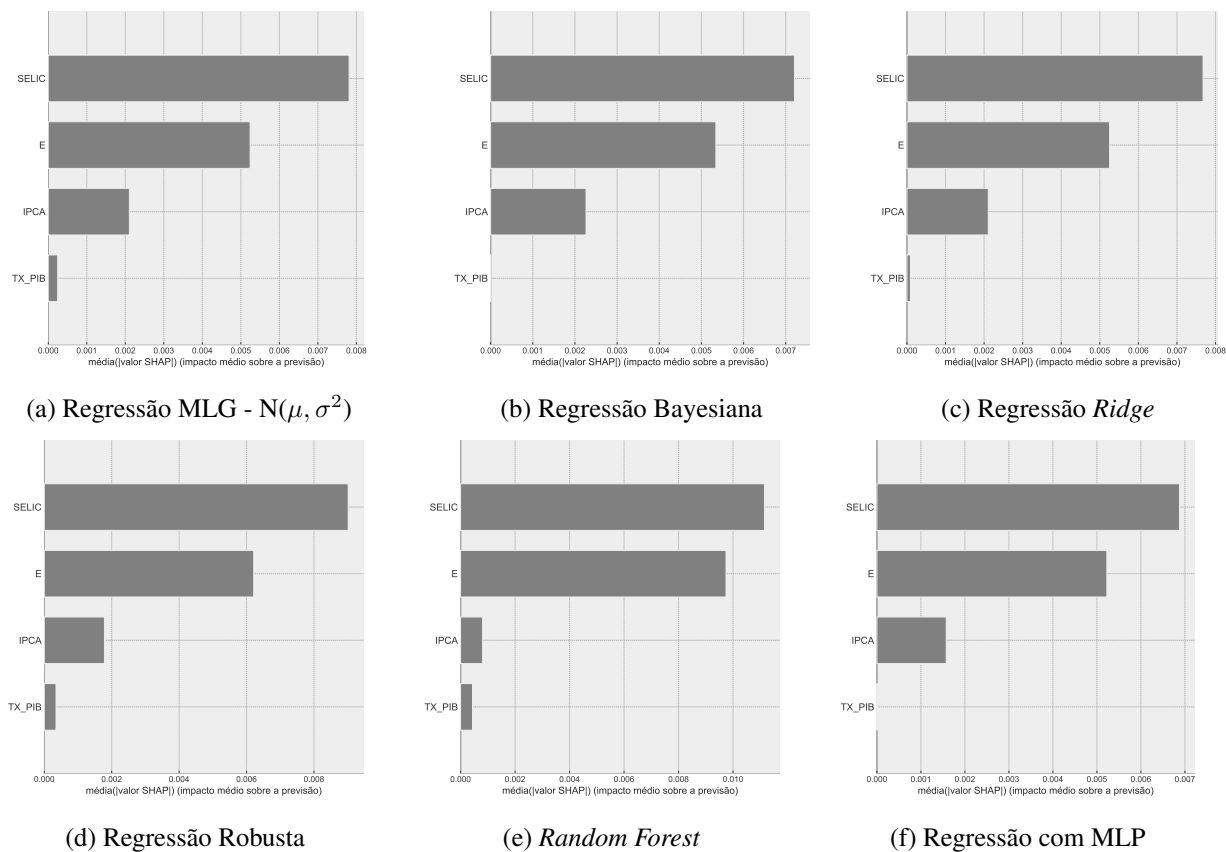


Figura 3: Impacto médio das variáveis selecionadas na previsão do grau de monetização da economia brasileira

## C Resultados para os Modelos Preditivos

Épocas	$\epsilon$	Nós na Camada Oculta	MSE	Modelo
120	0.010	5	0.058877	RNN
120	0.010	10	<b>0.007137</b>	RNN
120	0.010	20	0.011733	RNN
120	0.001	5	0.365443	RNN
120	0.001	10	0.022301	RNN
120	0.001	20	0.028734	RNN
120	0.010	5	0.045864	LSTM
120	0.010	10	0.030263	LSTM
120	0.010	20	0.022783	LSTM
120	0.001	5	0.224099	LSTM
120	0.001	10	0.160266	LSTM
120	0.001	20	0.110038	LSTM
120	0.010	5	0.009152	GRU
120	0.010	10	0.028267	GRU
120	0.010	20	0.021657	GRU
120	0.001	5	0.044227	GRU
120	0.001	10	0.151606	GRU
120	0.001	20	0.093635	GRU

Tabela 5: Resultado das simulações para previsão do grau de monetização da economia brasileira

## Referências

- S. Aras and P. J. Lisboa. Explainable inflation forecasts by machine learning models. *Expert Systems with Applications*, 207:117982, 2022. ISSN 0957-4174. doi: <https://doi.org/10.1016/j.eswa.2022.117982>. URL <https://www.sciencedirect.com/science/article/pii/S0957417422012106>.
- Banco Central do Brasil. Sistema gerenciador de séries temporais, 2021. URL <https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>. Online; aessado em 27 de fevereiro de 2021.
- C. M. Bishop and N. M. Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley, 2015.
- L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001. URL <https://link.springer.com/article/10.1023/a:1010933404324>.
- M. Buckmann, A. Joseph, and H. Robertson. An interpretable machine learning workflow with an application to economic forecasting. Technical report, Bank of England, 2022. URL [https://www.bankofengland.co.uk/working-paper/2022/an-interpretable-machine-learning-workflow-with-an-application-to-economic-](https://www.bankofengland.co.uk/working-paper/2022/an-interpretable-machine-learning-workflow-with-an-application-to-economic-forecasting)
- F. Carvalho, L. F. Paula, J. Sicsú, F. Souza, and R. Studart. *Economia monetária e financeira: teoria e política*. Elsevier Brasil, 2012.
- K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation, 2014. URL <https://arxiv.org/abs/1406.1078>.

- K. V. S. da Costa, F. L. C. da Silva, and J. da Silva Cordeiro Coelho. Forecasting quarterly brazilian gdp: Univariate models approach, 2020. URL <https://arxiv.org/abs/2010.13259>.
- F. Doshi-Velez and B. Kim. Towards a rigorous science of interpretable machine learning, 2017. URL <https://arxiv.org/abs/1702.08608>.
- S. Fatima and G. Hussain. Statistical models of kse100 index using hybrid financial systems. *Neurocomputing*, 71(13):2742 – 2746, 2008. ISSN 0925-2312. doi: <https://doi.org/10.1016/j.neucom.2007.11.044>. URL <http://www.sciencedirect.com/science/article/pii/S0925231208002373>. Artificial Neural Networks (ICANN 2006) / Engineering of Intelligent Systems (ICEIS 2006).
- A. Graves. *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer-Verlag Berlin Heidelberg, 2012. ISBN 978-3-642-24797-2. doi: <http://doi.org/10.1007/978-3-642-24797-2>.
- T. Hastie, R. Tibshirani, and J. Friedman. *Random Forests*, pages 587–604. Springer New York, New York, NY, 2009. ISBN 978-0-387-84858-7. doi: 10.1007/978-0-387-84858-7\_15. URL [https://doi.org/10.1007/978-0-387-84858-7\\_15](https://doi.org/10.1007/978-0-387-84858-7_15).
- J. Hicks. Critical essays in monetary theory. 1967.
- S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL <https://doi.org/10.1162/neco.1997.9.8.1735>.
- A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. doi: 10.1080/00401706.1970.10488634. URL <https://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634>.
- C. C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5 – 10, 1957. ISSN 0169-2070. doi: <https://doi.org/10.1016/j.ijforecast.2003.09.015>. URL <http://www.sciencedirect.com/science/article/pii/S0169207003001134>.
- M. Y. Hu, G. P. Zhang, C. X. Jiang, and B. E. Patuwo. A cross-validation analysis of neural network out-of-sample performance in exchange rate forecasting. *Decision Sciences*, 30(1):197–216, 1999. doi: <https://doi.org/10.1111/j.1540-5915.1999.tb01606.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.1999.tb01606.x>.
- D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014. URL <https://arxiv.org/abs/1412.6980>.
- F. Lazzeri. *Machine Learning for Time Series Forecasting with Python*. John Wiley & Sons, 2021. ISBN 978-1-119-68236-3.
- S. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions, 2017. URL <https://arxiv.org/abs/1705.07874>.
- S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee. Explainable ai for trees: From local explanations to global understanding, 2019.
- W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, Dec 1943. ISSN 1522-9602. doi: 10.1007/BF02478259. URL <https://doi.org/10.1007/BF02478259>.
- C. Molnar. *Interpretable Machine Learning*. 2021. URL <https://christophm.github.io/interpretable-ml-book/>.



- J. A. Nelder and R. W. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384, 1972. URL <https://doi.org/10.2307/2344614>.
- E. M. Ronchetti and P. J. Huber. *Robust statistics*. John Wiley & Sons, 2009.
- D. E. Rumelhart and J. L. McClelland. *Learning Internal Representations by Error Propagation*, pages 318–362. MIT Press, 1987. ISBN 9780262291408.
- L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953. ISSN 0027-8424. doi: 10.1073/pnas.39.10.1095. URL <https://www.pnas.org/content/39/10/1095>.
- M. H. Simonsen and R. P. Cysne. *Macroeconomia*. Atlas, 2009.
- M. Sundararajan and A. Najmi. The many shapley values for model explanation, 2020.
- D. West, S. Dellana, and J. Qian. Neural network ensemble strategies for financial decision applications. *Computers & Operations Research*, 32(10):2543 – 2559, 2005. ISSN 0305-0548. doi: <https://doi.org/10.1016/j.cor.2004.03.017>. URL <http://www.sciencedirect.com/science/article/pii/S0305054804000693>. Applications of Neural Networks.
- M. West and J. Harrison. *Bayesian forecasting and dynamic models*. Springer Science & Business Media, 2006.
- P. R. Winters. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3): 324–342, 1960. doi: 10.1287/mnsc.6.3.324. URL <https://doi.org/10.1287/mnsc.6.3.324>.
- L. Yu, S. Wang, and K. Lai. A novel nonlinear ensemble forecasting model incorporating glar and ann for foreign exchange rates. *Computers Operations Research*, 32(10):2523 – 2541, 2005. ISSN 0305-0548. doi: <https://doi.org/10.1016/j.cor.2004.06.024>. URL <http://www.sciencedirect.com/science/article/pii/S030505480400156X>. Applications of Neural Networks.
- G. Zhang and M. Y. Hu. Neural network forecasting of the british pound/us dollar exchange rate. *Omega*, 26(4):495 – 506, 1998. ISSN 0305-0483. doi: [https://doi.org/10.1016/S0305-0483\(98\)00003-6](https://doi.org/10.1016/S0305-0483(98)00003-6). URL <http://www.sciencedirect.com/science/article/pii/S0305048398000036>.
- 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. doi: 10.1057/palgrave.jors.2601133. URL <https://doi.org/10.1057/palgrave.jors.2601133>.