PREDICTING THE TERM STRUCTURE OF INTEREST RATE: A MACHINE LEARNING APPROACH[*]

PREVISÃO DA ESTRUTURA A TERMO DA TAXA DE JUROS: UMA ABORDAGEM COM MACHINE LEARNING

Natanael Colório Teixeira[**]
Nelson Seixas dos Santos[***]

ABSTRACT

Over time, several studies have attempted to predict the term structure of interest rates, each study has proposed a methodology for this purpose, but there is still no predominant method. In this article, we propose the use of machine learning models applied to a database of macroeconomic variables. The results show that machine learning models are capable of making predictions as expected, however, in the short term they still do not consistently outperform Random Walk, only as the horizon increases. Using the Diebold-Mariano test, it was not possible to find evidence indicating the superiority of machine learning models, and further studies are needed to advance on the subject.

Keywords: Machine Learning. Interest Rate. Forecasts. Term Structure.

RESUMO

Ao longo do tempo, diversos estudos tentaram prever a estrutura a termo das taxas de juros, cada estudo propôs uma metodologia para tal, mas ainda não há um método predominante. Neste artigo, propomos a utilização de modelos de machine learning aplicados a um banco de dados de variáveis macroeconômicas. Os resultados mostram que os modelos de machine learning são capazes de fazer previsões conforme o esperado, no entanto, no curto prazo ainda não superam consistentemente o Random Walk, apenas à medida que o horizonte aumenta. Utilizando o teste de Diebold-Mariano, não foi possível encontrar evidências que indiquem a superioridade dos modelos de machine learning, sendo necessários mais estudos para avançar no assunto.

Palavras-chave: Aprendizado de Máquina. Taxa de Juros. Previsão. Estrutura a Termo.

---

[*]Course Completion Work presented in 2024/1, to the Department of Economic Sciences and International Relations of the Faculty of Economic Sciences of the Federal University of Rio Grande do Sul (UFRGS), as a partial requirement for obtaining the title of Bachelor of Economic Sciences.

[**]Graduating in Economic Sciences from the Federal University of Rio Grande do Sul (natasmice@icloud.com).

[***]Advisor. PhD in Economic Theory from the Faculty of Economics, Administration and Accounting of the University of São Paulo (FEA/USP). Master in Economics from the Graduate School of Economics at Fundação Getulio Vargas (EPGE/FGV-RJ). Professor at the Department of Economics and International Relations at the Federal University of Rio Grande do Sul. (nelson.seixas@ufrgs.br).

1 INTRODUCTION

The term structure of the interest rate is a function that relates time and interest rates as defined by (Fisher, 1930) where he introduced his expectations hypothesis theory to explain the dynamics of the term structure. This model describes how the term structure of interest rates was formed by agents' preferences for more liquid assets and their expectation of future changes in interest rates. Investors would accept the risk of holding less liquid assets in their portfolios, as long as they could be remunerated at a higher rate in the future (liquidity premium), in such case the long-term interest rate would tend to be higher than the short-term rate. The setup of Fisher was purely deterministic though, which makes it hard to analyse the consequences of random shocks in most variables that affects the term structure.

On the other hand, the general model of asset pricing available around the 1930's, introduced by (Bachelier, 1900), have already described asset prices movement as a continuous-time stochastic process called Brownian motion. It is stated that the change value of an asset varies at each given moment and is independent of previous prices. This is the basic assertion on (Fama, 1970)'s theory of efficient markets, which states that it is not possible to predict the future based on past information, in particular, its random walk formulation. In an attempt to fully incorporate the term structure theory to the stochastic setup firmly established by (Black; Scholes, 1973), (Vasicek, 1977) formulated a continuous time stochastic model where the spot interest rate follows a diffusion process, the price of a discounted security depends on the spot rate over time and markets are efficient, concluding it was possible to forecast the curve using a mean-reverting stochastic model.

Nearly a decade later, (Nelson; Siegel, 1985) introduced a stochastic term structure expectations theory which was able to generate enough smoothness to reveal specific maturity patterns. The model was also capable of reflecting the primary shape of the term structure and not just an approximation providing a better setup for predicting the curve beyond the sampling range. In the following years, there were extensions to the Nelson-Siegel Model, such as (Svensson, 1995), which introduced a model that allowed capturing more points along the interest curve, making a more accurate model for analyzing and predicting the term structure of the interest rate.

One of the most famous models developed in recent years is the Diebold-Li model (Diebold; Li, 2006), since it allowed the incorporation of macroeconomic variables in addition to recognizing the existence of price arbitrage, something that was not clearly done by previous models. The results presented by Diebold and Li show that a good model can exist without assuming any price arbitrage, however the model did not outperform Random Walk in short-term forecast horizons, only beating it in long-term horizons.

(Pooter, 2007) deals with the uncertainty of the parameters in addition to the uncertainty of the model itself. The main result obtained was that predictive performance of the models depends on the time period analyzed, that is. models that rely heavily on the most recent observation outperform any other more sophisticated model on the short term, something that is not always true on the long term. Therefore, till this date we cannot point to a single model that is capable of accurately forecasting the term structure for all scenarios.

Our objective is to determine whether machine learning (ML) methods are able to outperform Random Walk (Bachelier, 1900) in all forecast horizons or are restricted to just one given horizon; our null hypothesis, thus, can be formulated as: both ML models and random walk have the same accuracy. Our approach consists of creating a macroeconomic database and feeding it to various models and determine their accuracy when compared to a Random Walk benchmark and if any could be a potential candidate for an "all horizons model".

The models tested were: Linear Regression, Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996), Support Vector Machines (Cortes; Vapnik, 1995) and Random Forest (Breiman, 2001).

This work is organized as follows. In section 2, it is presented a brief review of the models that were created throughout the literature to predict the term structure, going from more classic models to the most modern ones today. In section 3, it is shown the data that was collected to feed the tested models, some data statistics, as well as some graphs to illustrate the behavior of the variables over the period. In this same section it also presented the methods used to process the data, and the machine learning models used. In section 4, the results are presented alongside a comment on them. In section 5, a conclusion regarding the work.

2 A PIECE ON TERM STRUCTURE MODELLING

Below is an exposition of the Random Walk model, which will be used as a benchmark for this study, in addition to the two models that have been widely studied and that have managed to outperform the benchmark in some situations. The idea is to verify whether machine learning models can outperform the benchmark in situations where the other two did not.

2.1 RANDOM WALK

Brownian motion is a continuous-time stochastic process where the value of a variable changes unpredictably at each time and each next step is independent of the previous step as defined by (Bachelier, 1900). A discrete-time formulation of Brownian motion is often called Random Walk and is described below:

$$P_{t+1} = P_t + \epsilon_{t+1} \tag{1}$$

Where:

- P_t is the value of the variable in time t.
- P_{t+1} is the value of the variable in time $t+1$.
- ϵ_{t+1} is a random error term in time $t+1$

The work proposed by (Bachelier, 1900) is based on the assumption that it is not possible to predict future movements based on past information, thus consolidating a solid basis for the Efficient Markets Theory, formulated by (Fama, 1970).

2.2 NELSON-SIEGEL MODEL

The model created by (Nelson; Siegel, 1985) aimed to present a parsimonious way of describing the yield curve, using few parameters but, at the same time, providing a clear and precise structure, level, curvature and steepness of the yield curve. The proposed model was based on the expectations theory of (Fisher, 1930), it was able to reflect the primary shape of the curve rather than just an approximation, which resulted in the possibility of predicting the curve beyond the sampling range.

The Nelson-Siegel model is described by:

$$y(t) = \beta_0 + \beta_1 \left(\frac{1 - e^{-\lambda t}}{\lambda t} \right) + \beta_2 \left(\frac{1 - e^{-\lambda t}}{\lambda t} - e^{-\lambda t} \right) \tag{2}$$

Where:

- $y(t)$: Interest rate for a bond with maturity t.
- β_0: Long-term level of the yield curve.
- β_1: Inclination.
- β_2: Curvature.
- λ: Exponential decay rate.

2.3 DIEBOLD-LI MODEL

(Diebold; Li, 2006) took the Nelson-Siegel model and introduced a number of changes to it, more specifically in the exponential component structure. The model sought to maintain the essence of the Nelson-Siegel model, but adding a temporal component to the equation. The goal was to not only fit the curve to the present moment, but to provide good future predictions.

The Diebold-Li model is described by:

$$y_t(\tau) = \beta_{0t} + \beta_{1t}\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{2t}\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \tag{3}$$

Where:

- $y_t(\tau)$: Interest rate over time t for a security with maturity τ.
- β_{0t}: General level of the time yield curve t.
- β_{1t}: Inclination t.
- β_{2t}: Curvature t.
- λ: Exponential decay rate.

Furthermore, the authors proposed dynamic parameters to the temporal components using an auto-regressive (AR) models to describe the evolution of parameters over time.

In this way, it was possible to implement the Nelson-Siegel model with the insertion of dynamism in the parameters, which, according to the authors, would allow for good future predictions of the yield curve. The model ends up losing to Random Walk in short horizons, but proved to be superior when forecasting in the long term, (Diebold; Li, 2006).

3 DATA AND METHODS

This section aims to objectively present the data that were used to carry out the application of the methods, in addition to this, the methods used to carry out the study will also be presented.

3.1 DATA

First, we set the analysis period to encompass the last three Brazilian federal administrations, meaning that our sample ranges from 2010 to 2022. This choice aims to capture the latest trends in interest rates in Brazil, while also providing a large enough sample to perform statistical inference.

We use not only a metric for interest rates, but also the most common variables associated with macroeconomic performance, since any monetary variable must be understood in its macroeconomic environment. The variables were chosen based on the work of (Junior, 2021), they are assumed to have an influence on the term structure. Those are: the basic interest rate, the exchange rate, the consumer price index, and the credit default swap.

The basic interest rate was measured by the SELIC rate, which is the average of government bonds traded in the Special System for Settlement and Custody of the Brazil Central Bank (SELIC). We used the PTAX800 as a measure of the exchange rate. Both the SELIC and the PTAX800 were extracted from the time series system of the Brazil Central Bank. As for the inflation index, the Broad Consumer Price Index (IPCA) was used, which is used as a structure for Brazilian inflation targets. We extracted it from the website of the Brazilian Institute of Geography and Statistics (IBGE). The National Treasury Bill (LTN) is a public debt security issued by the Brazilian Government. It is a fixed-rate security that does not have coupon payments. We used data from the term structure of the LTN with maturities of 1, 3, 6 and 12 months to construct a proxy for the yield curve, since we did not obtain the data necessary to construct the term structure of the interest rate. To do so we collected data from the Brazilian Association of Financial and Capital Market Entities. Finally, the Brazilian 5-year Credit Default Swap (CDS), which is a financial instrument that functions as a contract against the risk of default of a given agent, traded on the over-the-counter market of the stock exchange, the data were taken from Investing website.

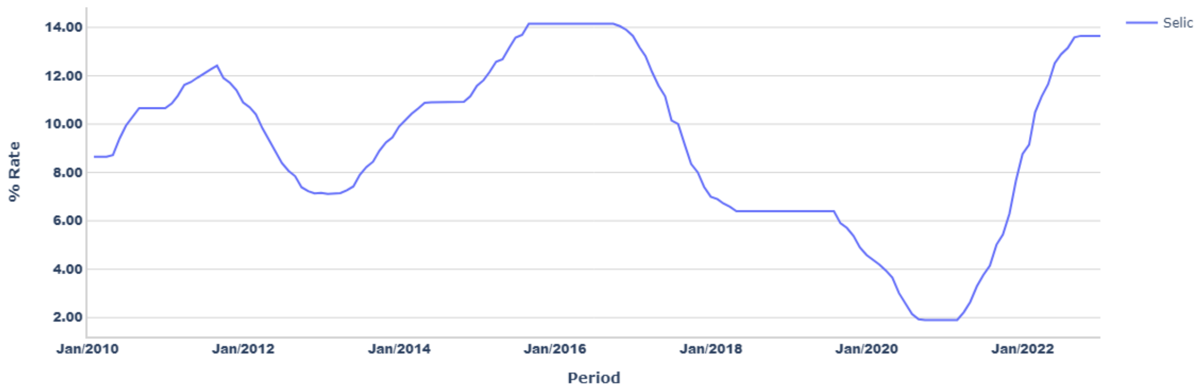
The data collected consist of daily observations (4185 observations). Since there are no trades on non-business days, missing data were disregarded.

In fact, SELIC rate, as shown in Figure 1, had a very tumultuous trajectory throughout the period analyzed from 2010 to 2012. The scenario was one of recovery after the 2008 financial Subprime Crisis when the Central Bank of Brazil maintained a policy of monetary tightening control inflation. SELIC rate went from 8.75% in 2010 to 12.5% in 2011 and was soon to 7.25% in 2012 in an attempt to stimulate economic growth.

In subsequent years, from 2013 to 2016, Brazil went through an economic crisis during former president Dilma Rouseff's government, with persistent inflation worsened by the political and economic crisis, so the Central Bank increased the rate aggressively to combat inflation. During this period, we went from a SELIC of 7.25% in 2013 to 14.25% in 2015, a level that remained until the end of 2016, resulting in a period of high inflation and economic recession.

After the impeachment of the then president Dilma, the country entered a slow economic recovery to control inflation. The SELIC, which was at 14.25%, was reduced to 6.5% in 2018. From 2019 onwards, the world faced the COVID-19 pandemic, resulting in a global economic crisis, the BCB reduced the SELIC rate to its historic low of 2% in 2020, from 2021 onwards, with the return of the upward trend in the inflation, the Central Bank resumed the upward cycle, raising the Selic rate again to levels that would exceed 11%.

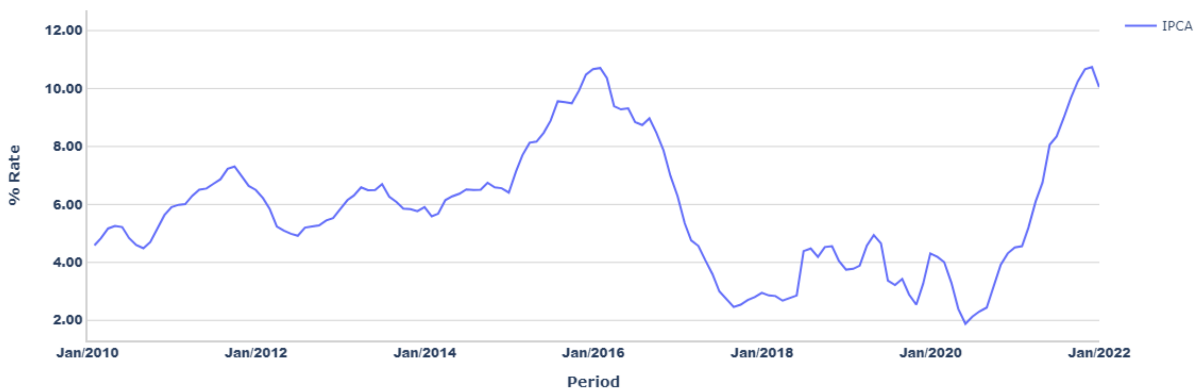
Figure 1 – Selic Rate



Note: Own work using data from Brazil Central Bank.

Just like the SELIC rate, in the years 2010 to 2012 the IPCA (Figure 2) also reacted to the recovery from the 2008 crisis, however it fluctuated within the inflation target, in the following years from 2013 to 2016, the IPCA ended up having a big jump, especially in the year 2015, the height of the economic crisis under Dilma Rouseff's government, reaching 10.67%. In the years 2017 to 2019, the IPCA came back under control despite the slow economic recovery, returning to high levels only in 2021, during the crisis generated by the COVID-19 pandemic, reaching around 10.06% in the same year.

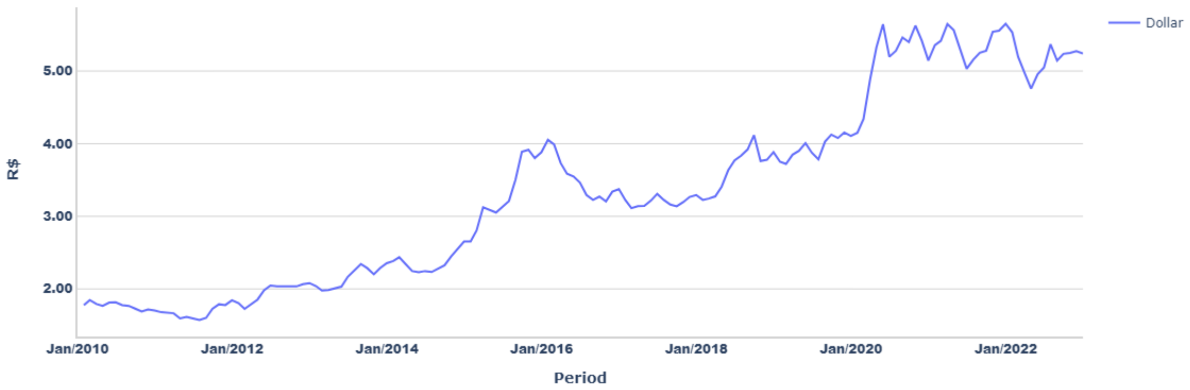
Figure 2 – IPCA



Note: Own work using data from IBGE.

The BRL/USD exchange rate (Figure 3) behaved similarly to the other variables, in 2010 it fluctuated between R$ 1.70 and R$ 1.80 and closed in 2012 at R$ 2.04. Again, during the economic crisis from 2013 to 2016, the Brazilian currency suffered a strong devaluation, the dollar continued on an upward trajectory, reaching a value of R$ 3.90 in 2015. In 2017 the dollar fell again due to economic policies. Even with the fiscal and economic recovery, the North American currency rose again due to the uncertainty caused by the Brazilian elections. Moving forward in time, to 2020, the dollar reached historic values, reaching a price of R$ 5.20, due to the COVID-19 pandemic that impacted the entire world, in 2021 the currency continued its upward trend, closing the year close to R$ 5.60.

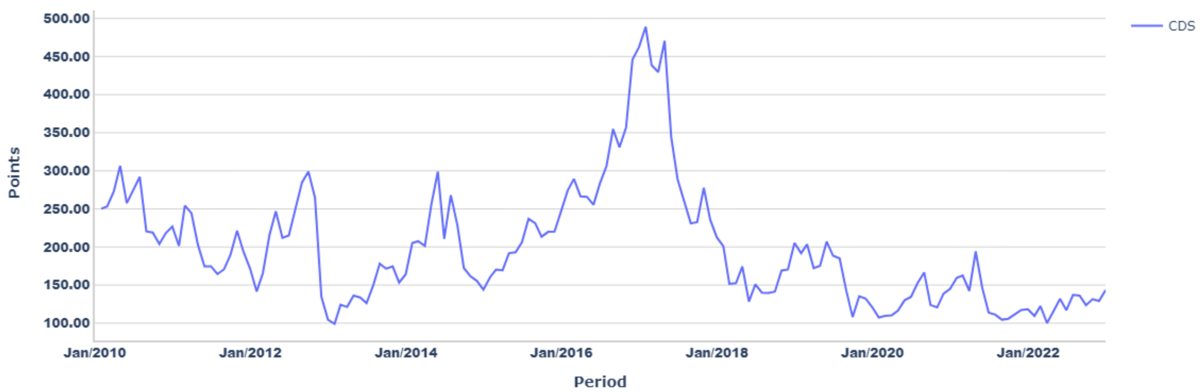
Figure 3 – Exchange



Note: Own work using data from Brazil Central Bank.

The CDS (Credit Default Swap) Figure 4 is a financial instrument that, in general terms, measures the credit risk of a given issuer, in short, it reflects the cost of protecting itself against the risk of default by an issuer. In Brazil, the CDS also went through fluctuations over time (Figure 5), in 2010 the CDS fluctuated between 100 and 150 basis points, rising to between 150 and 200 in 2011. In the period from 2013 to 2016, due to the economic crisis in Brazil, which increased distrust around the country, the CDS rose aggressively, surpassing 480 basis points in 2015. In the following years, the CDS returned to lower levels, remaining between 150 and 200 basis points in 2019. It rose again during the period of the COVID-19 pandemic, oscillating between 200 and 300 basis points, but did not reach levels as observed in 2015.

Figure 4 – CDS 5 Year



Note: Own work using data from Investing.

Below are presented the descriptive statistics of the variables. Table 1 and Figure

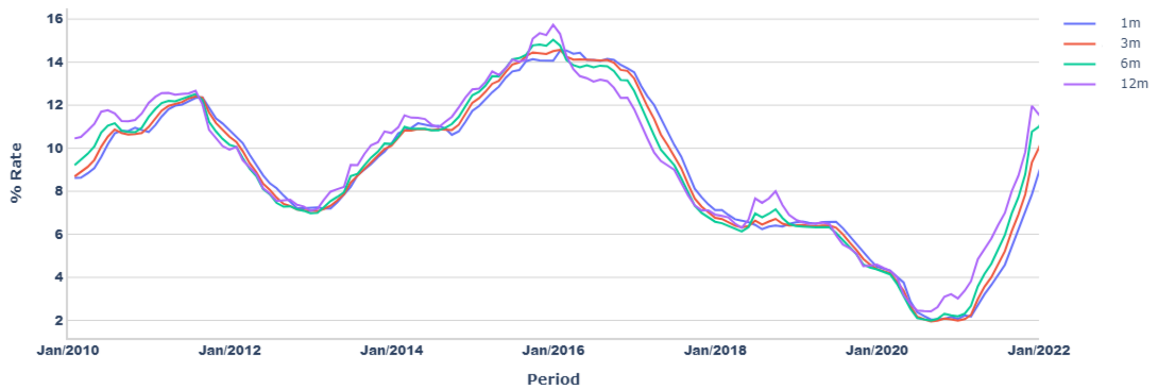
5 shows the data on the LTN term structure, used as a proxy for the yield curve. Table 2 shows the data on the features used to train the models.

Table 1 – Descriptive Statistics

	Mean	Std	Min	Max	25%	50%	75%
1m	9.207	3.464	1.914	14.884	6.588	9.636	11.981
3m	9.231	3.528	1.899	14.841	6.522	9.770	12.135
6m	9.313	3.533	1.875	15.408	6.613	9.878	12.281
12m	9.558	3.371	2.230	16.218	7.023	10.168	12.462

Note: Own work using data from Anbima.

Figure 5 – LTN Term structure



Note: Own work using data from Anbima.

Table 2 – Features Descriptive Statistics

	Mean	Std	Min	Max	25%	50%	75%
Selic (%)	9.1253	3.5292	1.8763	14.1579	6.4000	9.4000	11.9101
IPCA (%)	0.0001	0.0001	-0.0002	0.0005	0.00008	0.0001	0.0002
Dollar (R$)	3.3264	1.2930	1.5345	5.9372	2.0561	3.2369	4.1008
CDS (bps)	196.37	79.367	91.160	538.63	137.43	177.38	233.78

Note: Own work using data from Brazil Central Bank, IBGE and Investing website.

3.2 METHODS

The present study was carried out using machine learning, a subfield of artificial intelligence that aims to create a machine that, based on a certain set of data and answers, learn the rules that govern that system on its own. A regressor based in machine learning is, therefore, a machine that can estimate future events based on prior knowledge given by the user whithout explicitly being told what is the exact relation between those points of data, (Samuel, 1959).

Machine learning tries to simulate a human brain, and one of the essential parts of a functioning brain is learning. Today, the machine has shown that it can perform this type of task much faster than a human being could. This type of technology has opened up an unprecedented range of areas of knowledge, including economics.

For the development of this work, machine learning algorithms were used to make predictions. Based on the above, if we provide the machine with a sufficient quantity of data correlated to the variable we want to predict, the machine will be able to learn from this data and make predictions. In the following subsection it is explored the different types of algorithms that were used in this research.

Thinking about the nature of the data, as they are variables that have outliers, in addition to having different measurement units, data normalization was applied using the Robust Scaler (Pedregosa et al., 2011) algorithm. The algorithm was chosen because it is robust to outliers, thus preventing the models from being influenced by values and adding more precision to the models. Furthermore, preference was given to models that are resistant to overfitting, as presented below:

3.2.1 LINEAR REGRESSION

The first method applied is linear regression, which has the ability to make predictions by fitting a linear model to a set of data points given its target variable, the dependent and the independent, the dependent variables being the one that is the target of the prediction. Through this relationship, a linear equation is generated which shows how the values of the target variable tend to change as a result of the independent variables, (Rencher; Schaalje, 2008). Although the linear regression model depends on the linear independence between the variables, it is noted that the curves observed for the target are smooth, therefore, as there are no major differences between outliers, the model is valid for observation without representing a bias.

The model used aims to fit a linear model with coefficients

$$w = (w1, ..., wp) \tag{4}$$

in order to minimize the residual sum of squares between the variables observed in the database and the targets predicted by the linear approximation (Pedregosa et al., 2011). This linear approximation is described mathematically as:

$$\min_{w} ||Xw - y||_2^2 \tag{5}$$

3.2.2 LASSO MODEL

The LASSO model (least absolute shrinkage and selection operator) is a model that estimates sparse coefficients, using this method to select and regularize variables in order to increase the accuracy of forecasts. The model assumes that the coefficients of the linear model are, as mentioned above, sparse, which means that few of them are equal to zero (Tibshirani, 1996). The mathematical formulation behind the model used is given by:

$$\min_{w} \frac{1}{2n_{\text{samples}}} ||Xw - y||_2^2 + \alpha ||w||_1 \tag{6}$$

The LASSO model solves the minimization of the least squares penalty by adding $||w||_1$ where α is a constant and $||w||_1$ is the ℓ_1 norm of the vector of coefficients (Pedregosa et al., 2011).

3.2.3 SVM MODEL

The SVM model (support vector machines) is a model that classifies and regresses data. The aim of the model is to find a hyperplane that best separates the points between the target classes, this hyperplane is chosen in order to maximize the margin between the closest data points in different classes, and those points are given the name of support vectors, (Cortes; Vapnik, 1995). The model used is represented by the following equation:

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \max(0, 1 - y_i(w^T \phi(x_i) + b)), \tag{7}$$

3.2.4 RANDOM FOREST

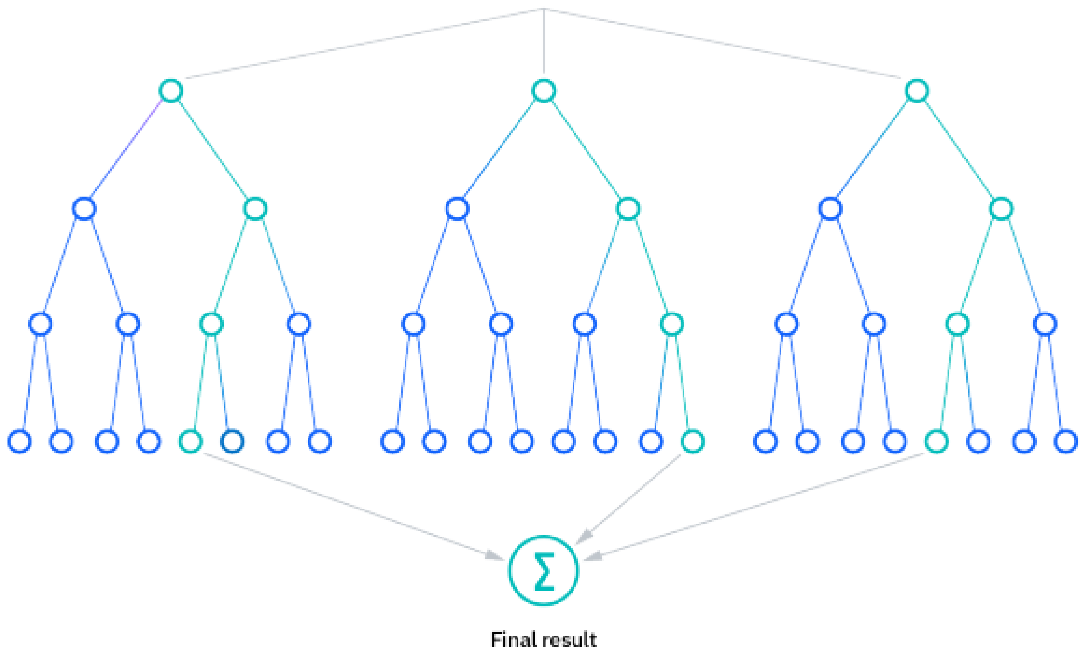
The Random Forest model (Breiman, 2001) is a machine learning model that uses decision trees that together work to increase the accuracy of the model. In this algorithm, each tree is an individual classifier, when the prediction of each tree is aggregated, it results in a more accurate output. A tree is a data structure that has nodes and branches, when this structure is complete, it is called a decision tree.

In this model there are so-called nodes and branches, each node itself represents a decision that is linked to an attribute of the data, while the branches represent the results of that decision. For each node, the model constructs a subset of attributes randomly. Attributes are nothing more than the data provided for training the models, in a straightforward way, they are features. The best division for the model is chosen through the generated subsets, instead of using all attributes, thus avoiding a larger domain of a specific tree. Additionally, the model trees are deepened as much as possible, to be defined by the user, to become highly specialized on this training data.

A problem that can be caused by the use of decision tree models is called overfitting, when the model fits very well to the training data, but ends up not being very effective for unobserved data. The Random Forest model, as explained above, separates the trees randomly using the attributes that were chosen for each subgroup, to perform the prediction, the model considers the attributes of each subgroup, which were chosen randomly, in this way, it ends up avoiding overfitting, because when a group of trees use the same attributes for their divisions, their correlation increases and they end up being induced to the same errors, with the random forest method, this problem is eliminated due to their division into subgroups with random attributes.

Below, you can analyze an example of what the structure of decision trees would be like in a random forest model:

Figure 6 – Random Forest Model



Note: Extracted from the IBM website.

3.2.5 CRITERION FOR METHOD COMPARISON

After making the forecasts, the results obtained for the maturities in each forecast horizon were compared with the reference Random Walk (Bachelier, 1900), which for the purpose of this study will be used as a benchmark because of its extensive use in literature as a benchmark for other models.

In order to determine the effectiveness of the models and make them comparable to each other, it is necessary to use a method that can capture the error between the observed variable and the predicted variable, for this study RMSE (Root Mean Squared Error)was chosen, which is also widely used in the literature to calculate the effectiveness of models (Arantes, 2013), the RMSE is described by the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{8}$$

The parameters are described as follows:

- $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ are the predicted values
- $y_1, y_2, \ldots, y_n$ are the observed values
- n is the number of observations

The RMSE method is a very objective way of verifying the adherence of models to reality, since it is calculated using the predicted rates and the true rates (Arantes, 2013).

Additionally, the Diebold-Mariano (Diebold; Mariano, 2002) test was performed to determine whether there is statistical relevance between the model predictions and the benchmark. A significance level of 5% was used, where a p-value less than 5% indicates that there is a significant difference between the benchmark and the model prediction, while a p-value greater than 5% indicates that there is no evidence to reject the null hypothesis.

The equation for the Diebold-Mariano test is presented below:

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2/T}} \tag{9}$$

Where:

- $\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$ is the average of the loss differences between the predictions of the two models, with $d_t = g(e_{1,t}) - g(e_{2,t})$ being the difference between the loss functions of the prediction errors $(e_{1,t})e(e_{2,t})$ of models 1 and 2 in time (t).

The variance $(\hat{\sigma}_d^2)$ is adjusted for autocorrelation and can be calculated as:

- $\hat{\sigma}_d^2 = \gamma_0 + 2\sum_{k=1}^{M} \gamma_k$ where (γ_k) are the autocovariances of the loss differences (d_t) in the gap (k) and (M) is the number of lags used in the adjustment.

The p-value can be calculated following the equation below, where CDF is the cumulative distribuition function and T refers to the degrees of freedom.

$$p\text{-value} = 2 \cdot (1 - CDF\left(|DM|, T-1\right))$$

3.2.6 THE SOFTWARE

To apply the methods described in subsection (3.2) we develop a application written in Python language (Foundation, 2023), helped by the following software libraries:

- Numpy - (Harris et al., 2020)
- Pandas - (McKinney, 2010)
- Scikit-Learn - (Pedregosa et al., 2011)
- Plotly - (Inc., 2015)
- Scipy - (Virtanen et al., 2020)

The source code is available on GitHub. The code was developed in an integrated development environment (IDE), using VS Code, developed by Microsoft Corp.

4 RESULTS AND DISCUSSION

Below will be presented the results obtained through the methods exposed in 3.2, in addition to discussing them, carrying out the analysis of the data resulting from the study. For the study, 80% of the database was used to train the model, predictions were made with out-of-sample data.

4.1 THE RESULTS

Following the methods presented in the previous section, tables 3, 4, 5 and 6 were constructed, showing the results obtained with the models presented above. The results show the scores obtained using the RMSE method, described in 3.2.5, where we can compare the accuracy of the machine learning models with the random walk benchmark. As described, horizons of 1, 3, 6 and 12 were used, the aim being to compare the effectiveness of the models in the short, medium and long term, so that it is possible to determine to what extent the models are effective and, if they fail, at what point this occurs.

The results can be seen below:

Table 3 – 1 month ahead forecast

	Random Walk	Linear Regression	LASSO	SVM	Random Forest
1m	0.737	0.339	2.960	0.341	0.606
3m	0.770	0.663	2.943	0.711	0.984
6m	0.828	1.090	3.034	1.174	1.327
12m	0.877	1.523	3.136	1.647	1.571

Note: RMSE for the 1-month horizon. Own Work.

Table 4 – 3 month ahead forecast

	Random Walk	Linear Regression	LASSO	SVM	Random Forest
1m	2.113	0.365	3.109	0.361	0.975
3m	2.190	0.758	3.080	0.794	1.338
6m	2.220	1.254	3.163	1.340	1.860
12m	2.078	1.749	3.242	1.912	2.266

Note: RMSE for the 3-month horizon. Own Work.

Table 5 – 6 month ahead forecast

	Random Walk	Linear Regression	LASSO	SVM	Random Forest
1m	4.191	0.408	3.279	0.401	1.507
3m	4.347	0.902	3.219	0.909	1.861
6m	4.365	1.508	3.271	1.556	2.408
12m	4.014	2.117	3.292	2.265	2.818

Note: RMSE for the 6-month horizon. Own Work.

Table 6 – 12 month ahead forecast

	Random Walk	Linear Regression	LASSO	SVM	Random Forest
1m	7.416	0.488	2.612	0.472	1.173
3m	7.867	1.176	2.469	1.147	1.427
6m	8.036	1.992	2.478	2.005	2.200
12m	7.525	2.821	2.375	2.874	2.460

Note: RMSE for the 12-month horizon. Own Work.

4.2 DISCUSSION

The tables above, 3, 4, 5 and 6, show the values obtained by the RMSE method to predict the term structure of interest rates in different maturities (1m, 3m, 6m and 12m) for different time horizons, comparing several machine learning models with the benchmark used, the random walk.

Analyzing the results, it is possible to see that the machine learning models cannot outperform the Random Walk in the short term. In some specific cases, the models manage to perform better, but this is not maintained constantly, as we can see in table 3.

Looking at the short-term horizon, which here will be the forecasts for 1 month and 3 months ahead, starting with the 1-month horizon, we can see that for the 1- and 3-month maturity, in some scenarios, some models outperform the benchmark, but for the 6- and 12-month maturities, all the machine learning models demonstrate a greater deterioration than the Random Walk. In the 3-month forecast horizon (table 4), we can a prevalence of machine learning models, with emphasis on SVM and Linear Regression models. In this horizon, the benchmark was surpassed by at least one machine learning model. However, other models were unable to surpass the benchmark, such as the LASSO model, which was inferior for all maturities in this horizon.

Comparing the two forecast horizons, 1 and 3 months, we can see that Random Walk is relatively superior to machine learning models 1 month in advance. However, it is defeated in some cases. However, there is no consistent superiority for this horizon. Furthermore, we can observe that as the horizon increases, the benchmark begins to deteriorate in a more notable way than other machine learning models.

Regarding the 6-month forecast (table 5) , once again the machine learning models demonstrate superiority over Random Walk. This time, all models outperform the benchmark, despite the deterioration of some of them. As we can see in Table 5, for all maturities, the Linear Regression and SVM models are superior to the benchmark and the other models. We can see that the Random Walk model deteriorates considerably as the forecast horizon continues to increase.

When it came to the 12-month horizon (table 6), the expectation was that the models would deteriorate considerably. However, the models managed to remain stable as the period increased, and the LASSO model even showed an improvement in relation to the previous period, demonstrating slightly better adaptability to longer horizons.

On the other hand, the Random Forest, Linear Regression and SVM models proved to be quite consistent, despite the horizon increasing and, therefore, the forecast becoming increasingly complex to execute. We can see that the random walk benchmark degrades more and more as the forecast horizon increases, which shows the failure of the model at long horizons compared to shorter horizons such as the 1-month horizon where it was quite accurate.

In the 12-month horizon, the longest, the superiority of machine learning models is evident, surpassing the benchmark for all observed horizons, in line with what was observed by (Diebold; Li, 2006), where its model, despite not being a machine learning model, began to demonstrate superiority over Random Walk only in longer horizons.

Machine learning methods have proven to be quite effective in predicting the time structure of interest rates. They have demonstrated an excellent ability to understand data and learn from it. However, despite being effective, they have not demonstrated the ability to consistently outperform Random Walk, outperforming it in longer horizons but being inferior in shorter horizons.

The objectives of this study were to verify the possibility of making time structure predictions using machine learning, and to verify whether any model could, on its own, outperform the benchmark in all cases. The first objective was met. We were able to demonstrate that the learning methods are quite consistent and accurate in making predictions. However, they show deterioration in some cases and end up falling behind the benchmark. Therefore, we cannot say that machine learning methods are superior to Random Walk for all horizons and maturities.

Furthermore, we were also able to verify that, although its predictive capacity is quite remarkable, a model alone is not capable of performing better in all horizons and for all time periods, in some cases a model was superior, but in others it ended up being surpassed by another model, therefore, the hypothesis that there could be a single machine learning model that was capable of surpassing the benchmark in all cases cannot be validated.

To validate the significance of the results of the machine learning models, the Diebold-Mariano (Diebold; Mariano, 2002) test was applied, which compares the results (tables 7, 8, 9 and 10) of each model with the Random Walk. The test compares the accuracy between two models to tell us if there is a statistically significant difference between the accuracy of the models. For this, the p-value was used. A result lower than 0.05 indicates that the result is significant, while higher values demonstrate that there is no evidence to support the superiority of the models.

The tables below show the p-values for each forecast horizon for each model compared to Random Walk:

Table 7 – 1 month ahead forecast

	Linear Regression	LASSO	SVM	Random Forest
1m	0.002	0.000	0.002	0.267
3m	0.076	0.000	0.216	0.001
6m	0.004	0.000	0.006	0.001
12m	0.001	0.000	0.018	0.000

Note: p-values for the 1-month horizon. Own Work.

Table 8 – 3 month ahead forecast

	Linear Regression	LASSO	SVM	Random Forest
1m	0.063	0.425	0.063	0.181
3m	0.071	0.457	0.069	0.209
6m	0.099	0.407	0.109	0.389
12m	0.299	0.307	0.602	0.478

Note: p-values for the 3-month horizon. Own Work.

Table 9 – 6 month ahead forecast

	Linear Regression	LASSO	SVM	Random Forest
1m	0.118	0.728	0.118	0.240
3m	0.118	0.656	0.117	0.251
6m	0.124	0.648	0.125	0.314
12m	0.137	0.743	0.146	0.416

Note: p-values for the 6-month horizon. Own Work.

Table 10 – 12 month ahead forecast

	Linear Regression	LASSO	SVM	Random Forest
1m	0.196	0.348	0.195	0.241
3m	0.139	0.223	0.138	0.167
6m	0.072	0.086	0.072	0.098
12m	0.005	0.000	0.004	0.003

Note: p-values for the 12-month horizon. Own Work.

We can see that for the 1-month horizon, the p-values indicate that there is a significant difference for most of the models tested, indicating that there is statistical relevance in the results. When we analyze the 3 to 12 months horizons (tables 8, 9 and 10), we can see that the test indicates that there is no evidence to support the hypothesis of superiority of the models in relation to the benchmark, thus, there is no evidence to reject the null hypothesis. Taking into account the results of the Diebold-Mariano test, we conclude that, although machine learning models are capable of making accurate predictions, it is not possible to state that these models are superior to Random Walk and more studies will be necessary to try to reject the null hypothesis.

5 CONCLUSION

The objective of this work was to determine whether it was possible to perform forecasts superior to Random Walk across all forecast horizons or whether machine learning methods would still be outperformed by the benchmark at a given time. To this end, a database of macroeconomic variables was built and applied to the models explained above. It was found that the machine learning models were able to make accurate predictions, but in the short term they were inferior to Random Walk, surpassing it in some cases, but not maintaining consistent superiority.

This brings us in line with the work carried out by (Pooter, 2007), where, through his study, he concludes that a forecasting model is not capable of producing good numbers alone in all horizons, therefore a combination of models is necessary according to the scenario so that we can have a good metric for the variable we want to predict (Pooter, 2007).

Bringing this context to the work carried out, we were able to realize that despite the good capacity of the machine learning models, none of them managed to make the best predictions for all horizons and all maturities. What we can observe is that, as the forecast horizon increases, the benchmark deteriorates more in relation to the other models.

In general, we can conclude this study with two considerations: a) The machine learning models tested here, with the incorporation of macroeconomic variables, are capable of making good forecasts of the target variable; b) The machine learning models tested are not capable of outperforming Random Walk for all forecast horizons, differing more significantly in the long term.

Furthermore, the Diebold-Mariano test was applied and the p-value calculated. The test results indicate that for the shorter horizon there is a statistical difference between the model results and the Random Walk results. However, for all other forecast horizons, the vast majority of results exceeded the value of 0.05 (5%), indicating that there is no evidence to reject the null hypothesis. Therefore, the final conclusion we can draw from this study is that machine learning models, despite being capable of making predictions, are still not superior to the Random Walk for all horizons.

Thus, the models tested do not differ much from models already existing in the literature, such as the Nelson-Siegel Model (Nelson; Siegel, 1985) and the Diebold-Li Model (Diebold; Li, 2006). At this point, further research is needed to test other models, in addition to trying to expose the models to other databases, to verify whether the models can improve their performance or whether they are really not superior enough to discard Random Walk.

REFERENCES

APLICADA, I. de P. E. **Ipeadata**. 2024. Available at: ⟨http://www.ipeadata.gov.br/Default.aspx⟩. Accessed on: Jun 20, 2024.

ARANTES, B. d. O. **Previsão da estrutura a termo da taxa de juros brasileira usando redes neurais artificiais**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2013. Available at: http://hdl.handle.net/10183/79126. Accessed on: Mar 22, 2024.

BACHELIER, L. **théorie de la spéculation**. In: Annales scientifiques de l'École normale supérieure. [S.l.: s.n.], 1900. v. 17, p. 21–86. Available at: http://www.numdam.org/item/10.24033/asens.476.pdf. Accessed on: Jun 5, 2024.

BANK, B. C. **Sistema Gerenciador de Séries Temporais**. 2024. Available at: ⟨https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries⟩. Accessed on: Jun 6, 2024.

BLACK, F.; SCHOLES, M. **The Pricing of Options and Corporate Liabilities**. Journal of Political Economy, The University of Chicago Press, v. 81, n. 3, p. 637–654, 1973. ISSN

00223808, 1537534X. Available at: http://hdl.handle.net/10183/79126. Accessed on: Apr 2, 2024.

BREIMAN, L. **Random forests**. Machine learning, Springer, v. 45, p. 5–32, 2001. Available at: https://link.springer.com/article/10.1023/a:1010933404324. Accessed on: Jul 15, 2024.

CORTES, C.; VAPNIK, V. **Support-vector networks**. Machine learning, Springer, v. 20, p. 273–297, 1995. Available at: https://link.springer.com/article/10.1007/Bf00994018. Accessed on: Jul 15, 2024.

DIEBOLD, F. X.; LI, C. **Forecasting the term structure of government bond yields**. Journal of Econometrics, v. 130, n. 2, p. 337–364, 2006. ISSN 0304-4076. Available at: https://www.sciencedirect.com/science/article/pii/S0304407605000795. Accessed on: Jul 16, 2024.

DIEBOLD, F. X.; MARIANO, R. S. **Comparing predictive accuracy**. Journal of Business & economic statistics, Taylor & Francis, v. 20, n. 1, p. 134–144, 2002. Available at: https://www.tandfonline.com/doi/abs/10.1198/073500102753410444. Accessed on: Jul 26, 2024.

FAMA, E. F. **Efficient capital markets**. Journal of finance, v. 25, n. 2, p. 383–417, 1970. Available at: https://www.degruyter.com/document/doi/10.7208/9780226426983-007/pdf?licenseType=restricted. Accessed on: Jun 24, 2024.

FISHER, I. **The theory of interest**. [S.l.]: New York: Macmillan. Kelley Publishers, 1930.

FOUNDATION, P. S. **"The Python Language Reference"**. [S.l.], 2023. Available at: https://docs.python.org/3/reference/index.html. Accessed on: Mar 3, 2024.

HARRIS, C. R. et al. **Array programming with NumPy**. Nature, Springer Science and Business Media LLC, v. 585, n. 7825, p. 357–362, set. 2020. Available at: https://doi.org/10.1038/s41586-020-2649-2. Accessed on: Mar 25, 2024.

IBM. **What Is Random Forest?** 2024. Available at: ⟨https://www.ibm.com/topics/random-forest⟩. Accessed on: Jun 6, 2024.

INC., P. T. **Collaborative data science**. Montreal, QC: Plotly Technologies Inc., 2015. Available at: https://plot.ly. Accessed on: Mar 31, 2024.

INVESTING. **CDS 5-Year**. 2024. Available at: ⟨https://br.investing.com/rates-bonds/brazil-cds-5-years-usd-historical-data⟩. Accessed on: Jun 19, 2024.

JUNIOR, P. I. F. d. S. **Estrutura a termo da taxa de juros no Brasil: projeções utilizando aprendizado de máquina**. Dissertação (Mestrado) — Universidade de Brasília, 2021. Available at: ⟨http://icts.unb.br/jspui/bitstream/10482/41996/1/2021_PedroIvoFerreiradeSouzaJunior.pdf⟩. Accessed on: Jun 4, 2024.

MCKINNEY Wes. **Data Structures for Statistical Computing in Python** . In: WALT Stéfan van der; MILLMAN Jarrod (Ed.). Proceedings of the 9th Python in Science Conference . [S.l.: s.n.], 2010. p. 56 – 61. Available at: https://pandas.pydata.org/docs/. Accessed on: Mar 28, 2024.

NELSON, C. R.; SIEGEL, A. F. **”Parsimoneous Modeling of Yield Curves for U.S. Treasury Bills”**. [S.l.], 1985. (Working Paper Series, 1594). Available at: http://www.nber.org/papers/w1594. Accessed on: May 18, 2024.

PEDREGOSA, F. et al. **Scikit-learn: Machine Learning in Python**. Journal of Machine Learning Research, v. 12, p. 2825–2830, 2011. Available at: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https:/. Accessed on: Jun 28, 2024.

POOTER, M. D. **Examining the Nelson-Siegel class of term structure models**. [S.l.], 2007. Available at: https://www.econstor.eu/handle/10419/86575. Accessed on: Mar 31, 2024.

RENCHER, A. C.; SCHAALJE, G. B. **Linear models in statistics**. [S.l.]: John Wiley & Sons, 2008.

SAMUEL, A. L. **Some studies in machine learning using the game of checkers**. IBM Journal of research and development, IBM, v. 3, n. 3, p. 210–229, 1959. Available at: https://ieeexplore.ieee.org/abstract/document/5391906. Accessed on: Jul 8, 2024.

SVENSSON, L. E. **Estimating forward interest rates with the extended Nelson & Siegel method**. Sveriges Riksbank Quarterly Review, v. 3, n. 1, p. 13–26, 1995. Available at: https://www.larseosvensson.se/files/papers/95QR.pdf. Accessed on: Jun 18, 2024.

TIBSHIRANI, R. **Regression shrinkage and selection via the lasso**. Journal of the Royal Statistical Society Series B: Statistical Methodology, Oxford University Press, v. 58, n. 1, p. 267–288, 1996. Available at: https://academic.oup.com/jrsssb/article/58/1/267/7027929?login=false. Accessed on: Jul 26, 2024.

VASICEK, O. **An equilibrium characterization of the term structure**. Journal of Financial Economics, v. 5, n. 2, p. 177–188, 1977. ISSN 0304-405X. Available at: https://www.sciencedirect.com/science/article/pii/0304405X77900162. Accessed on: May 26, 2024.

VIRTANEN, P. et al. **SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python**. Nature Methods, v. 17, p. 261–272, 2020. Available at: https://rdcu.be/b08Wh. Accessed on: Apr 1, 2024.

APÊNDICE A – FOLHA DE APROVAÇÃO

NATANAEL COLÓRIO TEIXEIRA

PREVISÃO DA ESTRUTURA A TERMO DA TAXA DE JUROS: UMA ABORDAGEM COM MACHINE LEARNING

Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Economia.

Aprovado em: Porto Alegre, ____ de _______ de 2024.

BANCA EXAMINADORA:

Prof. Dr. Nelson Seixas dos Santos – Orientador
UFRGS

Prof. Dr. Fernando Augusto Boeira Sabino da Silva
UFRGS

Prof. Dr. Ronald Otto Hillbrecht
UFRGS