Navigating Financial Uncertainty: Machine Learning Systemic Risk Detection in Brazil*

Gabriel Pires
UNESP

Diego M. Silva
Petros

April 3, 2025

Abstract

This paper aims to develop tools for anticipating systemic and abrupt fluctuations in asset prices, focusing on the Brazilian economy. While the literature on systemic risk has advanced significantly with the introduction of robust machine learning techniques offering superior predictive capabilities, most studies remain centered on the U.S. economy. To address this gap, we propose a combination of three machine learning models tailored to predict systemic events in Brazil, achieving an accuracy of 84%. The first model is a random forest algorithm that forecasts recessions based on macroeconomic indicators. The second and third models analyze asset price behaviour to identify potential systemic events, collectively enhancing predictive performance in a complex economic environment.

Keywords: Recession, economy, machine learning.

*We would like to thank Bruno Marchetto, Atila Riggo and Andre Luis Debiaso Rossi for their comments on the work carried out. We would also like to express our gratitude to the participants of the conferences where the work was presented, such as ABRAPP, ICDS and ANCEP. We would also like to thank Henrique Jäger and Petros for their support and encouragement of this research.

1 Introduction

Systemic risk refers to the potential collapse of an entire financial system or market, significantly impacting interest rates, exchange rates, and asset prices, thereby affecting the broader economy. Systemic crises can arise from natural disasters, such as pandemics, tsunamis, or earthquakes, or they may be triggered by financial factors, including market manipulation, labor strikes, energy price shocks, or bank failures[1]. Notably, systemic risk is not solely driven by external shocks. The normal functioning of markets can, over time, generate systemic fragility, necessitating the implementation of macroprudential regulatory measures. A well-known example is the work of Diamond and Dybvig (1983), which demonstrates a sunspot-driven bank run equilibrium.

A seminal paper by Burns (1946) introduced an approach to measuring market cycles, emphasizing the need to analyze multiple time series to better understand economic fluctuations. Later,Stock and Watson (1988) proposed a model for economic cycles, highlighting the role of unobservable factors in explaining these phenomena. Stock (1989) revisited economic indicators developed by Burns (1946) in collaboration with the NBER (National Bureau of Economic Research), leading to the creation of three indicators, including the recession index. More recently, Stock and Watson (2008) developed a recession probability model using coincident and leading variables. Addionally, Estrella and Mishkin (1998) was the first to incorporate variables such as interest rate, stock prices, and spreads to estimate recession probabilities.

Several studies have estimated recession probabilities based on the presence of unobserved common factors that can be extracted through linear combinations, such as trends, cycles, seasonality and volatility[2]. These studies use probit model to estimate recession probability. More recently, Costa et al. (2023) proposed a model using the framework of Issler and Vahid (2006) incorporating big data, as advocated by Stock and Watson (2014). The key advantage of big data is that it enables the model to provide real time recession predictions.

Given that systemic risk can emerge from a variety of sources, it is challenging to

[1]Bisias et al. (2012)

[2] Engle and Kozicki (1993), Vahid and Engle (1993), Engle and Issler (1995) and Issler and Vahid (2006)

define a fixed set of explanatory variables capable of reliably predicting abrupt systemic fluctuations in asset prices. To address this issue, researchers have tried to use variables that, on one side, are very sensitive of a wide range of events, and, on the other side, can be associated with observed situations of systemics collapses. This approach has facilitated the development of effective predictive models. For example, Liu and Moench (2016) proposed a model that predict recession in the US economy as a metric of systemic risk. In its proposal, the difference in interest rates (interest rate term structure spreads) is used, using a probit model to quantify the probability of a recession. Notably, this approach has gained widespread adoption and is currently employed by several financial institutions as a key metric for systemic risk management.

The development of statistical field and data science methods have provided new opportunities for construction models with a better potential to deal with a large range of variables - an essential feature when anticipating systemic price fluctuations. Several studies have applied machine learning techniques to address similar challenges using data from US and Europe (Wang et al. (2021); Vrontos et al. (2021); Nyman and Ormerod (2017)). However, research focused on the Brazilian economy remains scarse. In this context, our objective is to develop a robust methodology for measuring systemic risk in Brazil, leveraging recent modeling advancements while avoiding a mere replication of approaches designed for the U.S. economy.

The brazilian recession literature is extensive in articles referring to measuring banking systemic risk (Tabak et al. (2013); Guimarães et al. (2022); Capelletto and Corrar (2008); Datz (2002); Ng (2012)). As previously discussed, systemic risk arises from a wide range of situations. Thus, focusing on banking systemic risk is insufficient, as other macroeconomic variables are involved in such phenomenon and must be incorporated into the modeling approach.

In this context, the work of Chauvet and Morais (2010) discussed a time-varying autoregressive probit model to predict recessions in Brazil using economic data with different frequencies. However, the use of machine learning models applied to the Brazilian economy are essentially used for inflation forecast. For example, Garcia et al. (2017) and Araujo and Gaglianone (2023) focus on predicting the Consumer Price Index (CPI) for Brazil, detailing

the datasets used and their sources.

This paper aims to enhance tools for anticipating systemic and abrupt variations in asset prices by applying machine learning techniques to Brazilian economic data. Additionally, we propose a methodology for predicting recessions as early as possible by integrating two auxiliary indicators into our machine learning framework. We evaluate Logistic Regression, Decision Trees, Random Forest, and XGBoost models motivated by the work of Zhang et al. (2020); Leo et al. (2019) and Zhu (2020). Furthermore, we incorporate the Turbulence Index and the Absorption Ratio, as defined by Kritzman and Li (2010); Kritzman et al. (2010), to enhance predictive performance.

2 Methodology

2.1 Data

The data set consists of 60 different variables divided into four groups. The first group consists of price indexes and interest rates. The second aggregates information about the activity of the economy, such as the GDP and car production, for example. The third group consists of fiscal variables, such as debt as a percentage of GDP and in nominal terms. Finally, the last one is the set of financial data. All variables are presented in Table 5 in Appendix A. All information was obtained on the Brazilian Central Bank website[3] for the period between January 2002 and December 2024, on a monthly basis. Our approach for spiting the training and test set was made not only about data sample, but also thinking about the recessions periods. As recession isn't a common economic event we thought to include some period in the test set for evaluating the predictive power of machine learning models. So, the training set consists of data between January 2002 and December 2014, and the test set was between January 2015 and December 2024.

Some models are sensitive when data have different scales, negatively affecting modeling quality. One method to overcome this problem is the use of data normalization, a technique in which the aim is to standardize scales within a predicted range, transforming the data

[3] `https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries`

used into a standardized unit of measurement. Thinking about possible data leaks between the training and testing bases, the calculation was established on the training basis.

A recession for Brazil is defined as two negative quarters in a row. It should be noted that there may be divergences between recession dates from diverse sources, since the concept of recession may be different according to the author, and source. We used the definition of two negative quarter GDP for the sake of clarity and simplification, since other concepts tend to be a little more nuanced.

The vast majority of observations are for periods without recession. For the case of this study, less than 18% of the 276 data points were target labels for periods of true economic recession in Brazil. Because of that, the data were imbalanced, which can interfere with machine learning modeling, generating bias in its results. A key point is the evolution of tree models in machine learning are capable of dealing with imbalanced datasets. (Fernández et al. (2018))

2.2 Machine Learning approaches

The application of machine learning techniques in this context fits the problem we want to solve. This technique is very appropriate for solving problems that can be addressed via classification of results and also due to the vast capacity to deal with a large number of variables. In general, machine learning techniques for this specific area fall into classification techniques which have the ability to demonstrate the probability of a certain class of the model occurring or not. In general, given a set of training data $(x_1, y_1), ..., (x_n, y_n)$ models for this type of task can identify two classes: true (1) and false (0).[4]

$$f : R^D \rightarrow \{1, 0\} \tag{1}$$

To conclude which model presents the best performance for the proposed objective, the use of statistical metrics can be used to evaluate its performance. Normally, in order to improve performance, hyperparameter optimization techniques are used. The models evaluated in this work focused on different classes of machine learning, from the most classic

[4]See James (2013) and Deisenroth et al. (2020)

to Bagging and Boosting techniques. Four different techniques are discussed: Logistic Regression, Decision Tree, Random Forest, and XGBoost.

2.2.1 Logistic Regression

Logistic regression is a classic statistical technique and is appropriate when the dependent variable is dichotomous (binary) and can present independent variables in diverse ways. The data classification capacity refers to the sigmoid function used, as shown in Equation 2. (DeMaris (1995))

$$P(y = 1|x) = \frac{1}{1 + e^{-\beta_0+\beta_1 x}} \tag{2}$$

As the dependent variable of logistic regression is nonlinear, its cost function must be derived through the maximum likelihood estimate, which will be minimized.

$$\hat{\theta} = argmin\frac{1}{m}\sum_{i=1}^{\mathrm{m}} L(f(x^{(i)};\theta), y^{(i)}) \tag{3}$$

Where θ is the weight, L is the loss function, i is the data space for $(x^{(i)}, y^{(i)})$ and m the total number of instances. Finally, the classification between classes will be given by the probability resulting from the model, as shown below.

$$Decision(x) = \begin{cases} 1 & if\ \mathrm{P(y=1—x)>0.5} \\ 0 & \mathrm{otherwise} \end{cases} \tag{4}$$

Logistic regression is a widely used technique but has some negative points such as not solving problems that present non-linear relationships in addition to requiring that there is no collinearity between the data which, depending on the data used, may exist.

2.2.2 Decision Trees

Decision Trees are non-parametric supervised learning models used for both classification and regression tasks. These structures represent a hierarchical series of choices, in which each node in the tree corresponds to a question about the attributes of the data, and the

edges connect the answers to these questions. (Safavian and Landgrebe (1991); Swain and Hauska (1977))

The quality of candidates to be separated by each node is calculated using an impurity or loss function (H), this function being dependent on the task to be performed (regression or classification). For this work, the Gini criterion was used, described by Equation 5.

$$H(Q_m) = \sum_k p_{mk}(1 - p_{mk}) \tag{5}$$

Where p_{mk} is the resulting classification for a class k for a node m.

Decision trees are models with a certain robustness but can be very easy to overfitting the data, since only one tree is created to adjust (and learn) all the variables present in the modeling.

2.2.3 Random Forest

Random Forest is a machine learning algorithm based on decision trees. Its construction takes place through multiple independent decision trees that combine their predictions to obtain a more accurate and robust result. This procedure is feasible due to the use of techniques such as bootstrapping and bagging, which increase the diversity of the trees created. (Kullarni and Sinha (2013); Parmar et al. (2019); Shaik and Srinivasan (2019))

The bagging method was proposed by Breiman (1996) and presents itself as a solution to reduce the variance of a data set. To do this, a data set is generated by bootstrapping sampling of the original data.

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{\mathrm{B}} \hat{f}^{*b}(x) \tag{6}$$

The equation above represents the Bagging process where B represents the different sets generated by bootstrapping and $\hat{f}^{*b}(x)$ is the model to be trained for a given set. Then, the classification models are trained independently and, finally, a result is generated by aggregation, in the case of the classification task, this aggregation is voting.

Each tree will give a rating for each instance, votes from all trees are combined and counted. The class that presents the highest number of votes will be the classification

for the analyzed instance. Equation 7 describes the margin function which represents the confidence of tree classification.

$$margin(x, y) = av_k I(h_k(x) = y) - max_{j \neq y} av_k I(h_k(x) = j) \tag{7}$$

2.2.4 XGBoost

The XGBoost algorithm resembles that of Random Forest. In XGBoost, a series of decision trees are created in sequence, in which each tree tries to correct the errors of the previous tree (boosting), resulting in a model with high robustness and accuracy. Furthermore, regularization and automatic feature selection techniques are used. (Chen and Guestrin (2016))

Equation 8 describes the calculation established by each tree.

$$F_m(X) = F_{m-1}(X) + \alpha_m h_m(X, r_{m-1}) \tag{8}$$

Where α_m and r_m are the regularization terms and the residuals computed in the ith tree, h_m is the function trained to predict.

Just like the Random Forest model, XGBoost has the ability to deal with a wide range of variables and avoid overfitting due to the techniques used in its modeling, such as boosting.

2.3 Hyperparameter optimization

To optimize the models' hyperparameters, the GridSearch technique was used. This technique performs a complete search across the entire subset of hyperparameter space of the algorithm in question (Liashchynskyi and Liashchynskyi (2019)). the hyperparameter search space for logistic regression consisted of penalty, stopping criteria, and regularization strength. Basically, for tree models, they shared some search hyperparameter such as number of trees, maximum depth, minimum number of samples required to be at a leaf node, minimum number of samples required to split an internal node. To evaluate the performance of the models tested within the hyperparameter search space, the F1-Score metric was used, with the TimeSeriesSplit technique being used for cross-validation.

2.4 Evaluation Metrics

To choose the best model, error metrics related to supervised learning for classification models were evaluated, namely: Accuracy, Precision, Recall, F1-Score and ROC-AUC. The purpose of a classification model is to generate classes, in this case binary (True and False, 1 and 0, respectively). Knowing this, it is necessary to establish statistical methods capable of evaluating models and to do so, the classes generated by the model are evaluated. (Hand (2012))

As a result, these binary results can be in four ways:

- True Positive (TP): correct classification of the Positive class.
- False Negative (FN): error in which the model predicted the Negative class instead of Positive (type 2 error).
- False Positive (FP): error in which the model predicted the Positive class instead of Negative (type 1 error).
- True Negative (TN): correct classification of the Negative class.

Equation 9 describes the calculation of accuracy, which is one of the metrics capable of evaluating a classification model in general, quantifying how many classes the model got right. Accuracy quantifies the number of correct answers that the model obtained, for true and false, analyzed independently.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{9}$$

Equation 10 describes the calculation for quantifying precision for the true class. In other words, if the positive class is associated with the occurrence of a recession, the precision criterion sends the following message: within the set of times that the model indicated the occurrence of a recession, what is the proportion of times that this recession event fact occurred.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Recall quantifies the sensitivity of the model in classifying data to a given expected class. Equation 11 describes the calculation for quantifying sensitivity to the true class. In other words, and again assuming that the positive class indicates the occurrence of a recession, the message would be within the set of times in which a recession was observed, what is the proportion of times that the model previously indicated this occurrence.

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

Equation 12 describes the calculation of the F1-Score, which is the harmonic mean between Precision and Recall.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{12}$$

The Receiver Operating Characteristic (ROC) is also used in the evaluation in order to increase the probability to use the best model. The ROC is a curve that demonstrates how well a machine learning model can distinguish between two classes. For its determination, the true positive and false positive rates are used. A straight line is usually drawn on the diagonal of this graph, symbolizing a random classifier which would have a 50% chance of identifying the class of the analyzed instance.

2.5 Auxiliary Indicators

The papers of Kritzman and Li (2010) and Kritzman et al. (2010) define the Turbulence Index and the Absorption Ratio. The Turbulence Index refers to a situation in which asset prices, given their historical pattern of oscillation, start to show different behavior (extreme movements, decoupling of assets, convergence of uncorrelated assets, etc.). Equation 13 describes the mathematical expression for calculating the Turbulence Index, which was derived from the Mahalanobis Distance, and was previously discussed by Chow et al. (1999).

$$d_t = (y_t - \mu)\Sigma^{-1}(y_t - \mu)' \tag{13}$$

Where y_t is the vector of returns for period t, μ is the average of historical returns, Σ is the covariance matrix.

The Absorption Ratio seeks to identify periods in which the level of risk (volatility) between assets is coupled, signaling a situation of spread of volatility between different assets. Equation 14 describes its expression, which is modeled as the fraction of the total variance compared to the set of eigenvectors.

$$AR = \frac{\sum_{i=1}^{\mathrm{N}} \sigma_{E_i}^2}{\sum_{j=1}^{\mathrm{N}} \sigma_{A_j}^2} \tag{14}$$

Where N is the number of assets, $\sigma_{E_i}^2$ is the variance of the ith eigenvector, $\sigma_{A_j}^2$ is the variance of the jth asset.

3 Results and discussion

3.1 Machine Learning

The database was divided into in-sample and out-of sample groups for analysis, in which performance metrics were evaluated. Initially, the models were trained without any type of hyperparameter optimization and their performance metrics were computed. Then, hyperparameter optimization was performed using the GridSearch technique, and its metrics were measured again. The model with the best performance in both cases was the Random Forest model. The results are ilustrated in Table 1 and Table 2. The hyperparameter optimization technique, GridSearch, was carried out using the F1-Score as an evaluation metric, with 6 iterations and use of TimeSeriesSplit to create the cross-validation with 4 folds and no data shuffling.

Additionally to the random forest model, we conclude that XGBoost is also a suitable candidate. Logistic regression does not have many hyperparameters, because of that we have identical results with and without hyperparameter optimization. The results presented are in line with the advantages and disadvantages of each model discussed previously.

Logistic Regression presenting the worst performance due to its low capacity to deal with a large number of variables and the need for linear relationships between the data and the lack of multicollinearity Deisenroth et al. (2020). On the other hand, Random Forest

Models	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	37.38%	66.83%	54.11%	32.78%	64.95%
Decision Tree	71.92%	67.36%	62.17%	62.85%	62.17%
Random Forest	**82.24%**	**80.24%**	**75.56%**	**78.63%**	**81.24%**
XGBoost	72.90%	70.09%	72.28%	70.58%	80.34%

Table 1: Performance evaluation metrics of machine learning models (Logistic Regression, Decision Tree, Random Forest and XGBoost) before the hyperparameter optimization process using GridSearch.

Models	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	37.38%	66.83%	54.11%	32.78%	64.95%
Decision Tree	68.22%	66.53%	68.86%	66.39%	66.92%
Random Forest	**84.11%**	**81.49%**	**82.86%**	**82.08%**	**84.57%**
XGBoost	74.77%	71.40%	72.86%	71.93%	79.33%

Table 2: Performance evaluation metrics of machine learning models (Logistic Regression, Decision Tree, Random Forest and XGBoost) after the hyperparameter optimization process using GridSearch.

and XGBoost presented the best performances because they are robust models capable of dealing with different relationships between data and presenting techniques to avoid overfitting models such as bagging for Random Forest and boosting for XGBoost.

The ROC curves can be seen in Figure 1 and Figure 2. It is possible to observe the performance improvement of the Random Forest without hyperparameter optimization, since its curve tends to get closer to the 'perfect curve', which would be towards the value 1.0 of the TPR shaft. The Decision Tree and Logistic Regression curve are close to the "random" curve, which is the dashed line. This line is defined as random if the model had a 50% chance of classifying a given class, its curve would have this representation.

After training the Random Forest model, it was possible to obtain the recession probability graph (class 1) for Brazil, month/month, as shown in Figure 3. It is possible to observe the recession probability (black line), applied to moments of historical recession (hatched regions) beyond the model's decision threshold (probability of 0.5).

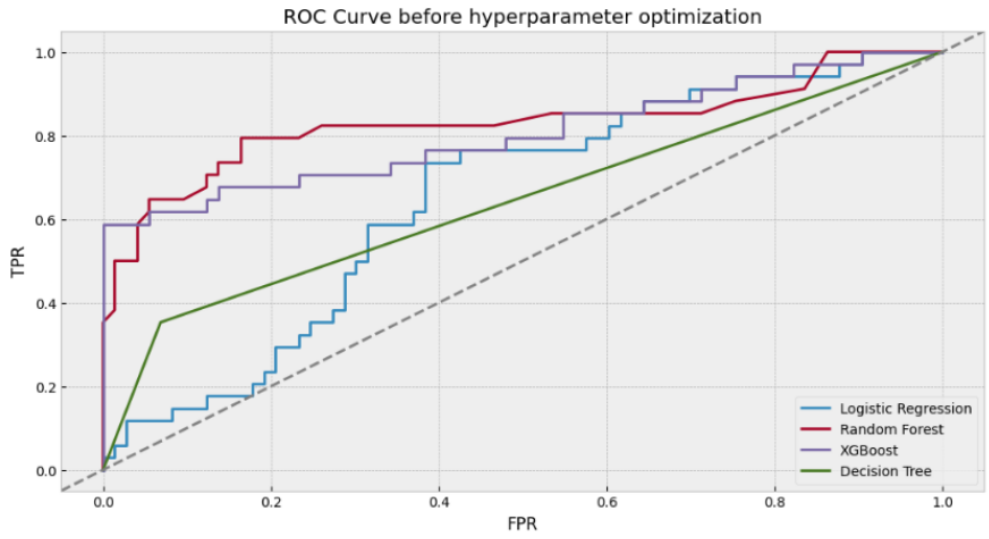

Figure 1: ROC curve for machine learning models before hyperparameter optimization. TPR: True Positive Rate; FPR: False Positive Rate.

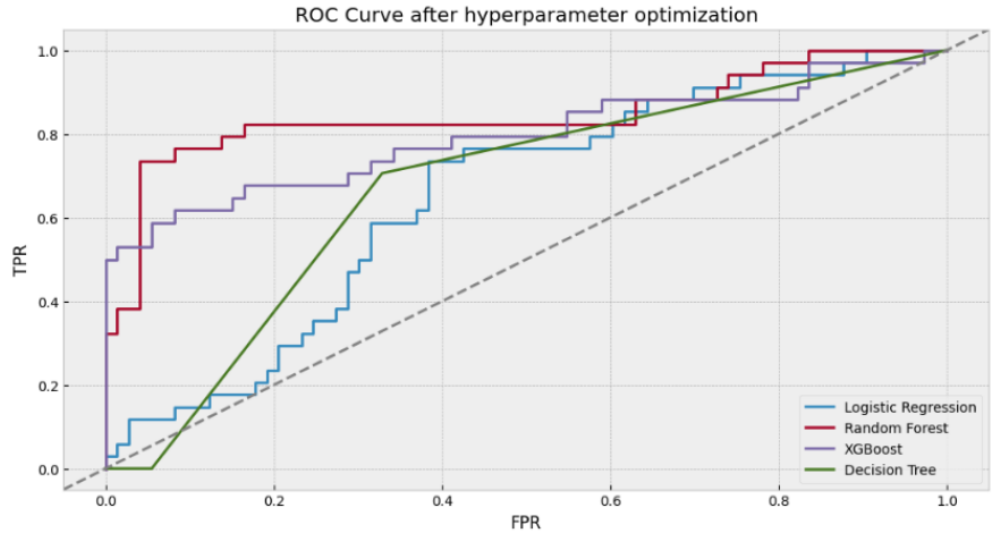

Figure 2: ROC curve for machine learning models after hyperparameter optimization. TPR: True Positive Rate; FPR: False Positive Rate.

The assertiveness of the model compared to history is notable, except in the period of COVID-19, due to the lack of explanatory data for the phenomenon that occurred: the pandemic was characterized as an economically exogenous event.

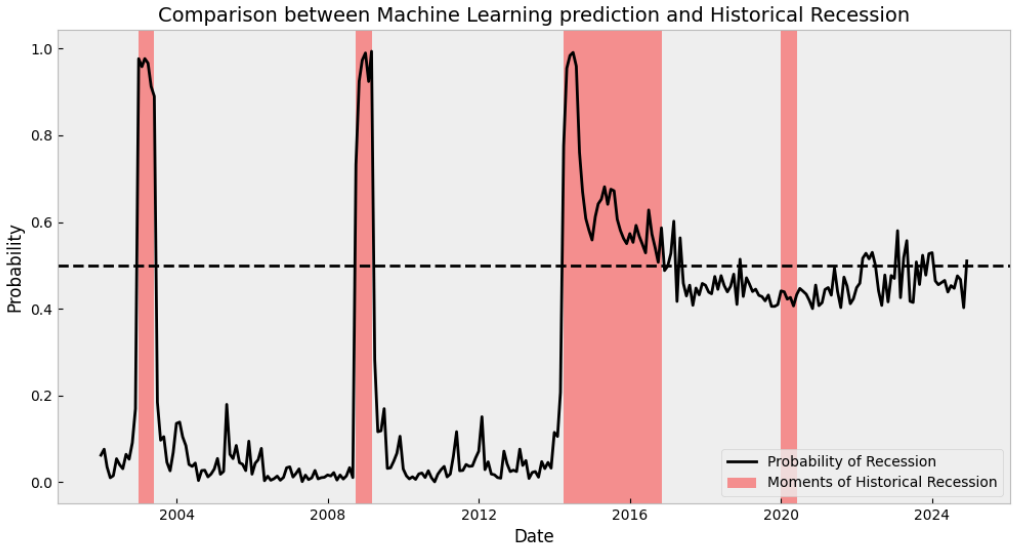

Figure 3: Probability resulting from the Random Forest model for predicting recession in the Brazilian economy.

3.2 Feature Importance

Models such as Random Forest were classified as black boxes due to their lack of interpretability. However, techniques have been developed to enable the analysis of variables that generally impact the decision-making of each node motivated by the specified metric, such as permutation importance.

3.2.1 Permutation Importance

Using the permutation importance technique, a random shuffle is performed between the data of one of the variables present in the model and then it is measured how much this modification would affect the predictive capacity of the model. The process occurs iteratively, but the calculation of feature importance can be represented by Equation 15. (Altmann et al. (2010))

$$i_j = s - \frac{1}{K}\sum_{k=1}^{\mathrm{K}} s_{kj} \tag{15}$$

Where s is the reference result for the model, K is the number of variables, s_{kj} is the

model result for the randomly shuffled variable.

The result of the importance permutation tells us how much the model's error will increase if that variable is randomly shuffled; the higher its value, the more importance that variable has for the model given the moment in time analyzed (new predictions can generate a new result, due to the dynamics of the new data used).

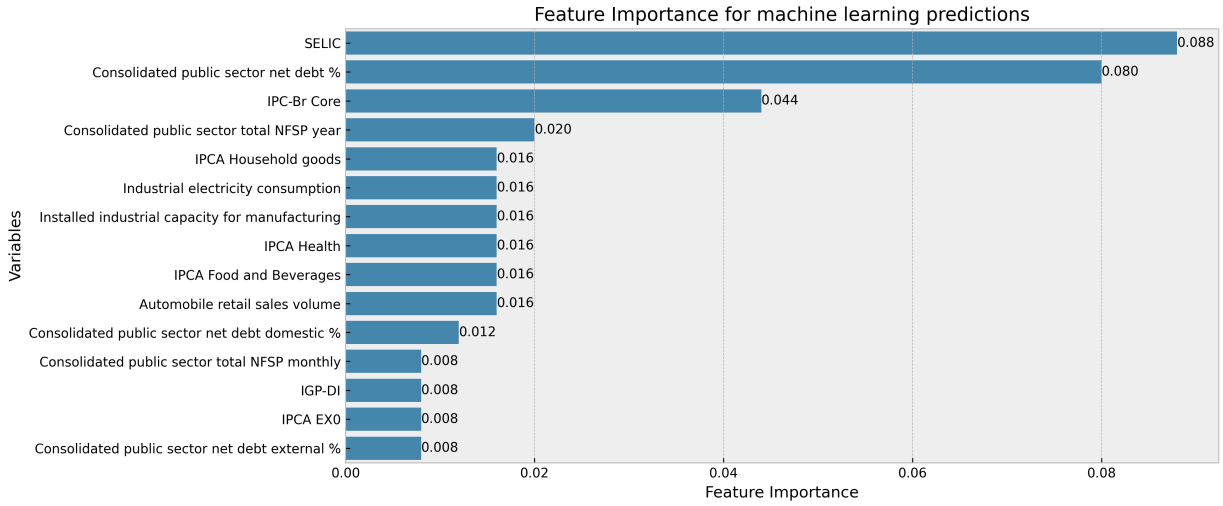

Figure 4: Result of the permutation importance analysis for the variables of the Random Forest model. Positive influences are show.

Therefore, for the prediction analyzed, "SELIC" and the "Consolidated public sector net debt %" presented greater importance in machine learning modeling, specific for the Random Forest, than the other variables to carry out the prediction.

3.3 Auxiliary Indicators

The random forest model performed well in classifying the period as either recession or non-recession, but it was unable to predict the recession. In this case, the model is not suitable to an risk manager or an investor. To address this issue, we incorporate the Turbulance Index and Absorption Ratio Models to our main model.

Our work uses Ibovespa, Selic, SP&500, Treasury, Commodities Index, and Real Estate Index as financial variables to incorporate the Turbulance Index and the Absorption Rate to the analysis.

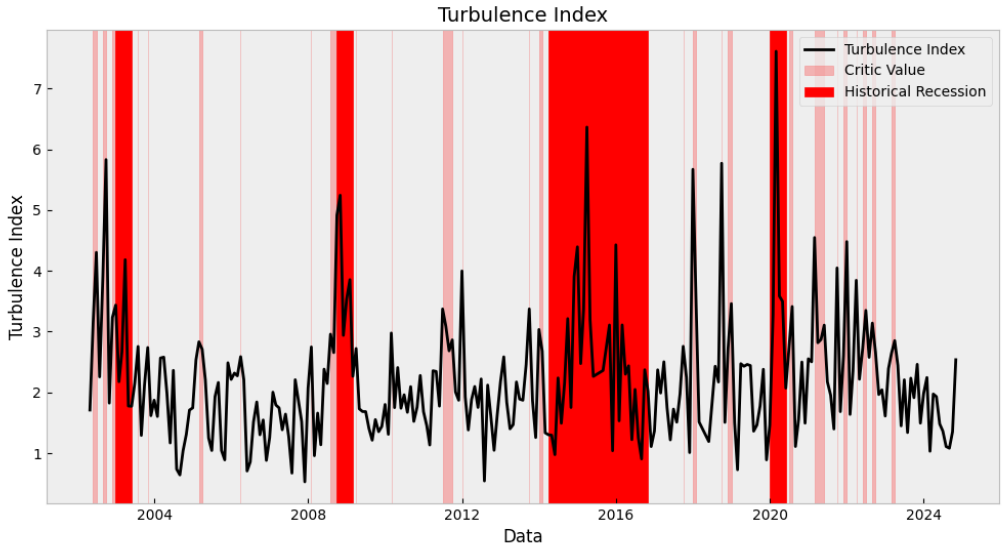

Figure 5: Result of modeling the Turbulence Index for the Brazilian economy.

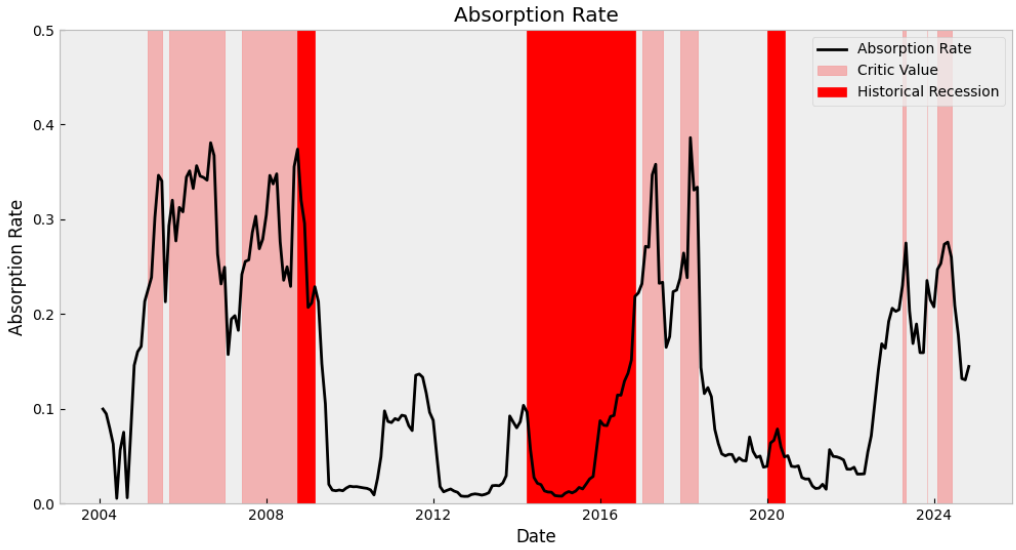

Figure 6: Result of the Absorption Ratio modeling for the Brazilian economy.

Initially we took them separately for analytical purposes, and both graphs (Figure 5 and Figure 6) show the critical values obtained through their models in dark hatched regions, the lighter hatched regions refer to moments of recessions in Brazil. According to the literature for both models, a threshold of 75% (values above the third quartile) was defined

to determine critical values, that is, values that indicate systemic risk.

It is possible to notice the presence of regions of critical values that precede periods of economic recession in Figure 5 and Figure 6. It is noteworthy that, among all historical periods of recession, the moment that preceded the 2009 recession was the one in which both indicators had the capacity to precede a recession 6 months in advance.

3.3.1 Comparison to VIX

As a way of validating the two indicators, a comparison was established with the VIX, a volatility index calculated using S&P 500 stock options.

In Figure 7, we can see a comparison between the Turbulence Index and the VIX. It is straightforward to notice that the indices are congruent, showing similar movements, including in amplitude. Figure 8 shows the comparison between the Absorption Rate and the VIX. In periods that preceded crises in the Brazilian economy, the absorption rate was comparable to the VIX, but with greater amplitudes, which helps in a clearer signaling of subsequent recessions.

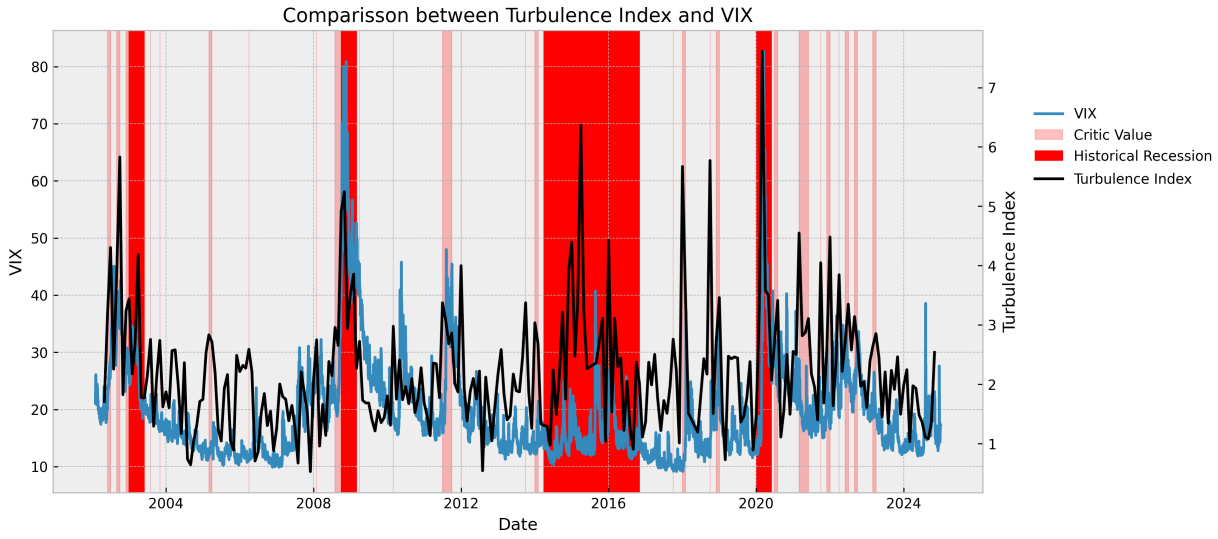

Figure 7: Comparison between the Turbulence Index model and VIX data.

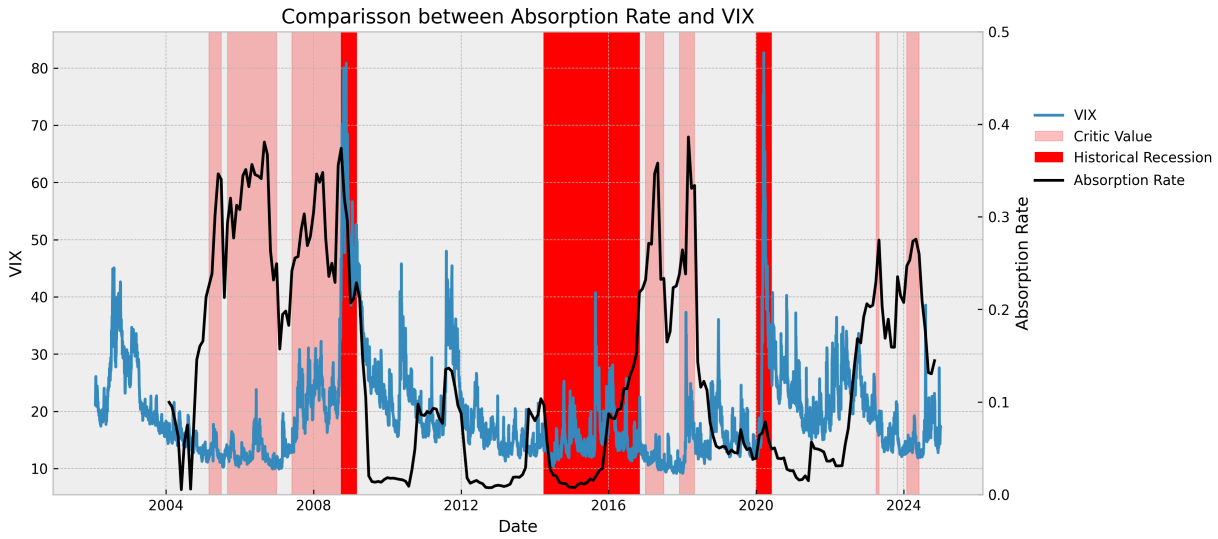

Figure 8: Comparison between the Absorption Ratio model and VIX data.

4 Final Model

The random forest model, illustrated in figure 3 was not able to detect the recession in 2008. However, the turbulance index, figure 5, and the absorption rate, figure 6, detected an abnormality in the analyzed pairs, which may indicate a situation of systemic risk for that period. It is worth returning to the main practical objective of developing a systemic risk prediction model: to prevent a possible situation of recession/systemic risk, through actions to reduce exposure to risk. In this sense, the auxiliary indicators seem to have a greater potential for anticipating recessions, while the model in figure 3 showed a greater capacity to validate a recession at the time of its onset. Observing the dynamics of the three graphs (Figure 3, Figure 5 and Figure 6) it can be noted that the last two generated warnings before the recession in 2003, 2009 and 2014. In this way, the three models considered signaled a high degree of complementarity.

We tested the hypothesis that the accessory indicators would have the capacity to assist in the result generated by the recession probability model through a weighted composition of its results, which took us to develop a final model with greater predictive potential. This weighting was developed using a data lag of 6 months before a period of recession, selecting three historical periods when the acessory indicators and the machine learning model predicted a recession period, in which we sought to solve a linear system of equations. The

goal is finding the multiples of these indicators which would result in a true classification, that is, recession. For this to happen, its result must be equal to or greater than 0.5.

$$0.5 \leq \begin{cases} \alpha_{ML}X + \beta_{TI}Y + \gamma_{AR}Z \\ \alpha_{ML}X + \beta_{TI}Y + \gamma_{AR}Z \\ \alpha_{ML}X + \beta_{TI}Y + \gamma_{AR}Z \end{cases} \tag{16}$$

Where $X = Y = Z = \{x | x \in R, 0 \leq x \leq 1\}$ are the probabilities found in the 6-month time window determined to carry out the system construction, α is the coefficient for the machine learning model, β is the coefficient for the turbulence index model and γ is the coefficient for the absorption ratio.

$$0.5 \leq \begin{cases} \alpha[\alpha v_k I(h_k(x_X) = y) - max_{j \neq y} a v_k I(h_k(x) = j)] + \beta[(y_t - \mu)\Sigma^{-1}(y_t - \mu)'] + \gamma[\frac{\sum_{i=1}^{N} \sigma_{E_i}^2}{\sum_{j=1}^{N} \sigma_{A_j}^2}] \\ \alpha[\alpha v_k I(h_k(x_X) = y) - max_{j \neq y} a v_k I(h_k(x) = j)] + \beta[(y_t - \mu)\Sigma^{-1}(y_t - \mu)'] + \gamma[\frac{\sum_{i=1}^{N} \sigma_{E_i}^2}{\sum_{j=1}^{N} \sigma_{A_j}^2}] \\ \alpha[\alpha v_k I(h_k(x_X) = y) - max_{j \neq y} a v_k I(h_k(x) = j)] + \beta[(y_t - \mu)\Sigma^{-1}(y_t - \mu)'] + \gamma[\frac{\sum_{i=1}^{N} \sigma_{E_i}^2}{\sum_{j=1}^{N} \sigma_{A_j}^2}] \end{cases} \tag{17}$$

We use 2009 recession to selected the weights due to the fact that both accessory indicators presented critical values before this period, but also generated a final model with a lower number of false positives.

The final model is composed of 39.3% of the Random Forest Model, 42.5% of the Turbulence Index and 18.2% of the Absorption Ratio. The final model is represented in Figure 9.

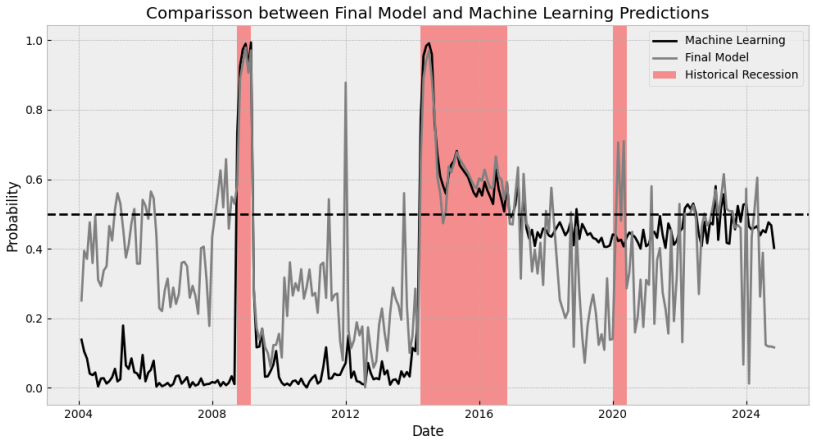

Figure 9: Comparison between the machine learning model, Random Forest, and the final model generated by weighting the three models presented for measuring systemic risk in the Brazilian economy: Random Forest, Turbulence Index and Absorption Ratio.

Looking at Figure 9, it can be seen that the Final Model had the ability to give signs of recession before its implementation. These events can be observed graphically before the crises in 2003, 2009 and 2014. Through Table 3, it is possible to observe, in bold, the moments in which the Final Model presented an indication of systemic risk before the crises highlighted above.

Date	Random Forest	Final Model	Recession
01/07/2002	0.0641	0.1760	0
01/08/2002	0.0312	0.4081	0
01/09/2002	0.0735	0.6139	0
01/10/2002	0.0795	0.1515	0
01/11/2002	0.1585	0.3282	0
01/12/2002	0.2385	0.2420	0
01/01/2003	**0.9166**	**0.9166**	**1**
01/07/2008	0.0147	**0.5501**	0
01/08/2008	0.0806	**0.5149**	0
01/09/2008	0.0494	**0.7594**	0
01/10/2008	**0.5839**	**0.5839**	**1**
01/11/2008	**0.8897**	**0.8897**	**1**
01/12/2008	**0.9334**	**0.9334**	**1**
01/01/2009	**0.9788**	**0.9788**	**1**
01/09/2013	0.0520	0.5681	0
01/10/2013	0.0544	0.2506	0
01/11/2013	0.0629	0.1763	0
01/12/2013	0.0788	0.1421	0
01/01/2014	0.2553	0.2886	0
01/02/2014	0.1818	0.0933	0
01/03/2014	0.2638	0.0941	0
01/04/2014	**0.6465**	**0.6465**	**1**

Table 3: Values found for the machine learning model, Random Forest, and the final model, weighting between the three models presented (Random Forest, Turbulence Index and Absorption Ratio) through the 6-month window before confirmation of a historic recession.

5 Conclusion

The recession probability model was developed using Random Forest, a supervised learning technique for classification tasks. In addition to generating a binary classification, the model also provides the probability of a given event occurring. The result present in figure 3 demonstrate a strong alignment between the model's predictions (periods in which the probability exceeds 0.5) and historical recession periods. The additional models, shown in figure 5 and figure 6, serve as early warning indicators for critical moments (potential recession or systemic risk events) by analyzing selected financial assets. This analysis essentially measures the impact of abnormal movements in one or more indices relative to the broader market. Notably, some of these identified critical moments precede historical recessions, reinforcing the model's potential as an early warning tool.

The combination of the three models tailored to predict systemic events in Brazil achieved an accuracy of 84%. To the best our knowledge, the literature on systemic risk in the Brazilian economy remains scarce and, to date, has not incorporated recent advancements in statistical and data science methodologies.

The proposed model demonstrates robustness and achieves satisfactory performance metrics for a classification model based on economic variables. Additionally, the probabilistic output of the model has proven effective in identifying recessionary periods, highlighting its potential as a valuable tool for systemic risk assessment in Brazil.

References

Altmann, A., L. Toloşi, O. Sander, and T. Lengauer (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics 26*(10), 1340–1347.

Araujo, G. S. and W. P. Gaglianone (2023). Machine learning methods for inflation forecasting in brazil: New contenders versus classical models. *Latin American Journal of Central Banking 4*(2), 100087.

Bisias, D., M. Flood, A. W. Lo, and S. Valavanis (2012). A survey of systemic risk analytics. *Annu. Rev. Financ. Econ. 4*(1), 255–296.

Breiman, L. (1996). Bagging predictors. *Machine learning 24*, 123–140.

Burns, A. (1946). Measuring business cycles. *National Ber-eau of Economics Research*.

Capelletto, L. R. and L. J. Corrar (2008). Índices de risco sistêmico para o setor bancário. *Revista Contabilidade & Finanças 19*, 6–18.

Chauvet, M. and I. Morais (2010). Predicting recessions in brazil. In *6th Colloquium on modern tools for business cycle analysis: the lessons from global economic crisis*.

Chen, T. and C. Guestrin (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Chow, G., E. Jacquier, M. Kritzman, and K. Lowry (1999). Optimal portfolios in good times and bad. *Financial Analysts Journal 55*(3), 65–73.

Costa, A. B. R., P. C. G. Ferreira, W. Piazza, O. T. C. G. Gaglianone, J. V. Issler, A. Brasil, and F. Rodrigues (2023). Predicting recessions in (almost) real time in a big-data setting. Technical report.

Datz, M. D. X. d. S. (2002). *Risco sistêmico e regulação bancária no Brasil*. Ph. D. thesis.

Deisenroth, M. P., A. A. Faisal, and C. S. Ong (2020). *Mathematics for machine learning*. Cambridge University Press.

DeMaris, A. (1995). A tutorial in logistic regression. *Journal of Marriage and the Family*, 956–968.

Diamond, D. W. and P. H. Dybvig (1983). Bank runs, deposit insurance, and liquidity. *Journal of political economy 91*(3), 401–419.

Engle, R. F. and J. V. Issler (1995). Estimating common sectoral cycles. *Journal of Monetary Economics 35*(1), 83–113.

Engle, R. F. and S. Kozicki (1993). Testing for common features. *Journal of Business & Economic Statistics 11*(4), 369–380.

Estrella, A. and F. S. Mishkin (1998). Predicting us recessions: Financial variables as leading indicators. *Review of Economics and Statistics 80*(1), 45–61.

Fernández, A., S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera (2018). *Learning from imbalanced data sets*, Volume 10. Springer.

Garcia, M. G., M. C. Medeiros, and G. F. Vasconcelos (2017). Real-time inflation forecasting with high-dimensional models: The case of brazil. *International Journal of Forecasting 33*(3), 679–693.

Guimarães, L., M. Castro, L. P. Silva, and A. Ugolini (2022). Fatores determinantes do risco sistêmico bancário brasileiro: Uma abordagem covar-cópula. *Brazilian Review of Finance 20*(4), 137–167.

Hand, D. J. (2012). Assessing the performance of classification methods. *International Statistical Review 80*(3), 400–414.

Issler, J. V. and F. Vahid (2006). The missing link: Using the nber recession indicator to construct coincident and leading indices of economic activity. *Journal of Econometrics 132*(1), 281–303.

James, G. (2013). An introduction to statistical learning.

Kritzman, M. and Y. Li (2010). Skulls, financial turbulence, and risk management. *Financial Analysts Journal 66*(5), 30–41.

Kritzman, M., Y. Li, S. Page, and R. Rigobon (2010). Principal components as a measure of systemic risk. *Available at SSRN 1582687*.

Kullarni, V. Y. and P. K. Sinha (2013). Random forest classifier: a survey and future research directions. *Int. J. Adv. Comput 36*(1), 1144–1156.

Leo, M., S. Sharma, and K. Maddulety (2019). Machine learning in banking risk management: A literature review. *Risks 7*(1), 29.

Liashchynskyi, P. and P. Liashchynskyi (2019). Grid search, random search, genetic algorithm: a big comparison for nas. *arXiv preprint arXiv:1912.06059*.

Liu, W. and E. Moench (2016). What predicts us recessions? *International Journal of Forecasting 32*(4), 1138–1150.

Ng, A. (2012). Avaliação do risco sistêmico dos bancos no brasil.

Nyman, R. and P. Ormerod (2017). Predicting economic recessions using machine learning algorithms. *arXiv preprint arXiv:1701.01428*.

Parmar, A., R. Katariya, and V. Patel (2019). A review on random forest: An ensemble classifier. In *International conference on intelligent data communication technologies and internet of things (ICICI) 2018*, pp. 758–763. Springer.

Safavian, S. R. and D. Landgrebe (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics 21*(3), 660–674.

Shaik, A. B. and S. Srinivasan (2019). A brief survey on random forest ensembles in classification model. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, Volume 2*, pp. 253–260. Springer.

Stock, J. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual 4*.

Stock, J. H. and M. W. Watson (1988). A probability model of the coincident economic indicators.

Stock, J. H. and M. W. Watson (2008). *Business cycles, indicators, and forecasting*, Volume 28. University of Chicago Press.

Stock, J. H. and M. W. Watson (2014). Estimating turning points using large data sets. *Journal of Econometrics 178*, 368–381.

Swain, P. H. and H. Hauska (1977). The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics 15*(3), 142–147.

Tabak, B. M., S. R. Souza, and S. M. Guerra (2013). Assessing systemic risk in the brazilian interbank market. *Banco Central do Brasil Working Papers* (318), 89.

Vahid, F. and R. F. Engle (1993). Common trends and common cycles. *Journal of Applied Econometrics*, 341–360.

Vrontos, S. D., J. Galakis, and I. D. Vrontos (2021). Modeling and predicting us recessions using machine learning techniques. *International Journal of Forecasting 37*(2), 647–671.

Wang, Z., K. Li, S. Q. Xia, and H. Liu (2021). Economic recession prediction using deep neural network. *arXiv preprint arXiv:2107.10980*.

Zhang, R., C. Yi, and Y. Chen (2020). Explainable machine learning for regime-based asset allocation. In *2020 IEEE International Conference on Big Data (Big Data)*, pp. 5480–5485. IEEE.

Zhu, F. (2020). Market regime forecasting using correlation networks and machine learning classifiers.

Appendix A

Activity	Inflation and Interest Rate	Fiscal	Financial
PIB (R$)	TJLP	Minimum wage	Ibovespa
Vehicle sales	IPC-Br	NFSP total monthly	Savings deposit
Total Energy Consumption	IGP-M	consolidated public sector	
Monthly Loans	IPA-DI	NFSP total consolidated public sector year	
Installed capacity	IPCA	NFSP total consolidated public sector year % GDP	
Brazil Metal Commodities Index	IPCA Food and Beverages	Total consolidated public sector net debt	
Car production	IPCA Housing	Internal consolidated public sector net debt	
Truck production	IPCA Residential items	External consolidated public sector net debt	
Bus production	IPCA Clothing	Total consolidated public sector net debt %	
Retail sales volume	IPCA Transport	Internal consolidated public sector net debt %	
Retail sales volume of fuels and lubricants	IPCA Communication	External consolidated public sector net debt %	
Sales volume retail supermarket, prod. Food	IPCA Health		
Sales volume retail fabric	IPCA Expenses		
Sales volume retail furniture electro	IPCA Education		
Auto retail sales volume	IPCA 15		
Sell cars	IPC-FIPE		
Commercial electrical energy consumption	IGP-DI		
Residential electrical energy consumption	IGP-10		
Industrial electrical energy consumption	INPC		
Other electrical energy consumption	Monetary base		
	SELIC		
	CDI		
	NCC		
	IPC-Br Core		
	IPCA EX0		
	IPCA EX1		
	IPCA double weighting		
	Smoothed trimmed mean IPCA		

Table 4: Variables used to create machine learning models.