# Ensenble of machine learning applied to economic cycles analysis: a comparative study using antecedent macroeconomic indicators for Brazilian GDP classification and prediction

Eduardo Palhares Júnior[a,b,*], Flavio A. M. Cipparrone[b]

[a]*Federal Institute of Amazonas, Itacoatiara 69101-030, Brazil*
[b]*Escola Politécnica, University of São Paulo, São Paulo 05424-970, Brazil*

---

**Abstract**

This work proposes a comparative study between several machine learning techniques, applied in the analysis of the phases of the Brazilian economic cycle. To this end, several macroeconomic indicators were used to build a model that was able to identify and predict the turning points of the economic cycle, such as the beginning of a recession or a recovery. The discretization of the variables proved to be decisive in the quality of the classification process, due to the diversity of the data and the non-linear nature of the analyzed phenomenon. The different techniques used reinforce a dilemma, because usually the best results come from very abstract methods, making it difficult to interpret the steps and their causes.

*Keywords:* Machine Learning, Classification, Economic Cycle

---

## 1. Introduction

Expert systems applied to the economy study are used to assist researcher and professionals in decision making regarding the forecast of the stock market, securities and commodities. This task was performed following with two

*Corresponding author
*Email addresses:* eduardo.palharesjr@ifam.edu.br (Eduardo Palhares Júnior), flavio@lps.usp.br (Flavio A. M. Cipparrone)

main approaches. One of the approaches explores the existing standards and relationships in the data without explicit reference to financial theories (Chang et al., 2009; Kamo and Dagli, 2009; Guresen et al., 2011; Svalina et al., 2013; de A. Araújo et al., 2015; Cervelló-Royo et al., 2015). On the other hand, there is another approach that explicitly considers theory in relation to market characteristics (Vasicek, 1977; Cox et al., 1985; ESTRELLA and HARDOUVELIS, 1991; Svensson, 1994; Estrella and Mishkin, 1997, 1998; Diebold and Li, 2006).

Understanding the dynamics and interrelation of macroeconomic variables with the future evolution of the economic cycle is a problem that arouses great interest to both academics and market participants. In the literature, we can identify two main currents of research. The theoretical current tries to explain the characteristics that determine the turning points of the cycle, through the behavior of the interest curve such as slope, level and curvature, based on several financial theories (Hicks et al., 1975; Vasicek, 1977; COX et al., 1981; BROWN and DYBVIG, 1986; HO and LEE, 1986; Nelson and Siegel, 1987; Heath et al., 1992; Svensson, 1994; Alexander et al., 2001). On the other hand, the empirical current seeks to implement the use of different economic methods and knowledge discovery techniques, with the aim of modeling the structure and dynamics of economic cycles (Ju et al., 1997; Kim and Noh, 1997; Zimmermann et al., 2002; Enke and Thawornwong, 2005; Jacovides, 2008; Chionis et al., 2009; Oh and Han, 2000; Vela, 2013; Gogas et al., 2014).

Recently, the use of knowledge discovery techniques has become increasingly common, mainly because they are capable of capturing and dealing with non -linearities between variables, as well as the complexity involved in seasonality and structural ruptures (Ju et al., 1997; Kim and Noh, 1997; Zimmermann et al., 2002; Jacovides, 2008; Oh and Han, 2000; Vela, 2013; Gogas et al., 2014). However, there are many difficulties involved in using this type of tool, as non -linearities considered in each stage of prediction are often extremely complicated, making interpretation unfeasible. Thus, many of these methods, despite having excellent results, are limited to being considered black boxes.

The use of knowledge discovery techniques discussed in the literature is carried out taking into consideration specific definitions of the problem, very peculiar characteristics of data sets and specific tools. Understanding the relationships of macroeconomic variables under the context of the behavior of economic cycles would greatly benefit from the use of a methodological approach, as usual in other areas of knowledge. This approach deals with the definition of a process that includes the standard steps considered in the decision making carried out by experts to solve analogous problems, allowing the comparison between different techniques, and finally making it possible to evaluate progress in relation to common objectives.

The classification of economic variables has been used for a long time as an instrument to predict relevant movements within economics studies, such as Burns and Mitchell (1946) who studied economic slowdown and recessions. Over time, several different variables have been proposed and evaluated as economic indicators, as discussed in Estrella and Mishkin (1995), which are notably recognized as important in identifying a point of inflection in economic growth movements. (Kauppi and Saikkonen, 2008; Rudebusch and Williams, 2009)

There are also several professionals and academics who concentrate efforts to produce analysis of the state of the economy, such as the Federal Bank of Chicago's national activity index CFNAI (2020), which consists of quantifying the weighted average of the 85 monthly indicators of state economic activity United. The idea behind this approach is that there are factors that are common to all of them, that is, using as an index can be a useful tool for predicting the stages of the economic cycle. In the results presented in Stock and Watson (1999) it was shown that CFNAI has a significant indicator about current economic activity. However, most economists follow a wide variety of indicators, and so far none of them have been totally reliable in the past, that is, they cannot be determined to evaluate the future state of the economy.

Thus, this article aims to develop machine learning based models that analyze various economic indicators, seeking a methodology that signals the possibility of a point of inversion of economic growth, in this case, the beginning of a recession phase in several temporal horizons.

## 2. Methodology overview

The general objective of this methodology is the study characterization of the behavior of economic cycles, based on statistical and machine learning techniques. Specifically, the context chosen for this study was the Brazilian macroeconomic scenario, through a model that describes GDP behavior as a function of other economic variables.

This work considers the definition of "economic cycle" as the cyclical process by which the economy floats between the stages of expansion and contraction. In each economic cycle, two stages or "inflection points", ie growth and recession, are defined in each economic cycle. However, economic theories also defined alternative characterizations of internships, such as the economic cycle of three stages (recovery, recession and stagnation), or, such as the economic cycle of four stages (initial recession, total recession, initial recovery, late recovery) (Bhaumin, 2011 [Online].).

It is important to emphasize that the choice of GDP as a variable of interest occurred due to the world socioeconomic context resulting from the Covid-19 pandemic, given the risk of widespread economic recession. However, the model is quite flexible and allows broader analysis of other variables of interest, simply changing the research hypotheses.

To facilitate continuous analysis and improvement of the project, the steps were modularized to facilitate the clearance of internships and understanding of the intermediate phenomena, as shown in the figure 1.

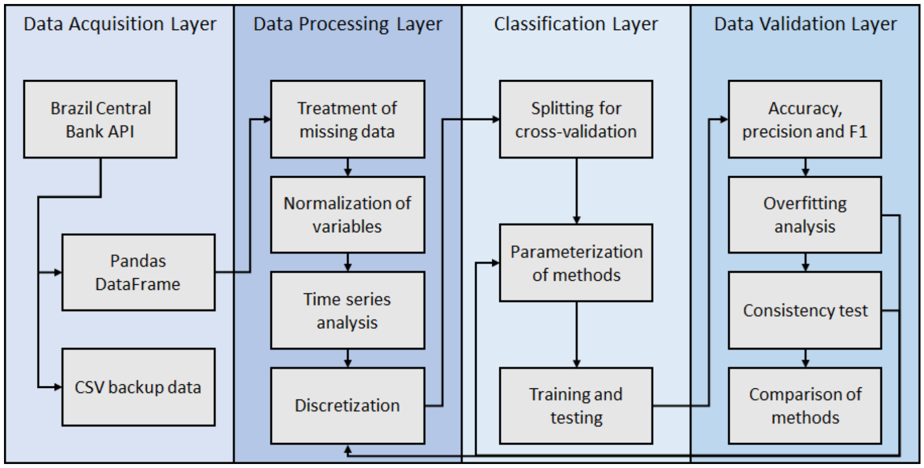


Figure 1: Architeture of methodology

### 2.1. Data acquisition and understanding

The first step involves understanding the requirements and the search for data available for analysis. In addition, it involves the collection and prior analysis of data consistency, comparing a theoretical perspective with a descriptive statistical analysis.

In this study, the data set used refers to macroeconomic data from the Brazil's economy according table 1, during the period between January 2002 and March 2021. The data were obtained directly from the Central Bank of Brazil, through an API provided by the financial authority from the country.

Table 1: List and description of the economic variables

| Economic Variable | Description |
|---|---|
| PIB | GDP monthly |
| IPA | Wholesale Price Index-Market |
| IPEM | Physical Production - Mineral extraction |
| IPIT | Physical Production - Capital goods |
| IPBC | Physical Production - Intermediate goods |
| IPBCD | Physical Production - Durable goods |
| IVVV | Sales volume index in the retail sector - Vehicles and motorcycles, spare parts - Brazil |
| VVCCL | Sales of factory authorized vehicle outlets - Light commercial cars sales |
| VVCC | Sales of factory authorized vehicle outlets - Trucks sales |
| IEF | Future expectations index |
| ICC | Consumer confidence index |
| Spub | Credit operations outstanding of financial institutions under public control - Total |
| Spriv | Credit operations outstanding of financial institutions under private control - Total |
| M1 | Money supply - M1 (working day balance average) |
| M2 | Broad money supply - M2 (end-of-periodo balance) |

Detailed analysis of descriptive statistics presented in Table 2 is an important component of the data understanding process, allowing us to identify stylized facts of the variables of interest. For example, the average and standard deviation of the variables provide an estimate of the long -term equilibrium value and their medium historical fluctuations that can be used as a reference for future interpretations of models.

Table 2: Summary statistics of the economic variables

| Economic Variable | Min | Median | Max | Range | Mean | Std dev | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| PIB | -0,11 | 0,01 | 0,10 | 0,22 | 0,01 | 0,04 | -0,15 | 0,01 |
| IPA | -0,02 | 0,01 | 0,07 | 0,09 | 0,01 | 0,01 | 1,37 | 3,84 |
| IPEM | -0,23 | 0,00 | 0,18 | 0,41 | 0,00 | 0,06 | -0,10 | 0,86 |
| IPIT | -0,25 | 0,00 | 0,21 | 0,46 | 0,00 | 0,07 | 0,00 | 0,61 |
| IPBC | -0,46 | 0,01 | 0,40 | 0,86 | 0,01 | 0,11 | -0,25 | 1,79 |
| IPBCD | -0,81 | 0,02 | 1,10 | 1,91 | 0,02 | 0,16 | 1,15 | 11,31 |
| IVVV | -0,44 | 0,01 | 0,51 | 0,95 | 0,01 | 0,12 | 0,42 | 1,83 |
| VVCCL | -0,52 | 0,02 | 0,63 | 1,15 | 0,02 | 0,16 | 0,08 | 1,12 |
| VVCC | -0,44 | 0,00 | 0,85 | 1,29 | 0,02 | 0,17 | 1,03 | 3,83 |
| IEF | -0,13 | 0,00 | 0,12 | 0,26 | 0,00 | 0,05 | 0,08 | 0,34 |
| ICC | -0,14 | 0,00 | 0,15 | 0,29 | 0,00 | 0,05 | 0,02 | 0,74 |
| Spub | -0,01 | 0,01 | 0,08 | 0,09 | 0,01 | 0,01 | 1,12 | 3,22 |
| Spriv | -0,02 | 0,01 | 0,05 | 0,07 | 0,01 | 0,01 | 0,28 | 0,32 |
| M1 | -0,09 | 0,01 | 0,12 | 0,22 | 0,01 | 0,02 | 0,98 | 9,77 |
| M2 | -0,01 | 0,01 | 0,06 | 0,07 | 0,01 | 0,01 | 1,89 | 5,72 |

The backup of the data set is very important, since the authority responsible for making available time series eventually modifies the collection/processing methodology of some variable, creating a new series and discontinuing the previous one. In the development and validation stage of the architecture this risk should be avoided as it is a source of unnecessary noise, so all the work was developed with the same database stored in the backup.

### 2.2. Data preparation

The second layer deals with the preparation of initial gross data, in which the variables are transformed and modified to properly feed the modeling step. This is an iterative process that assists in improving the analysis (modeling step) and may involve the addition or removal of variables or lags, missing data treatment, or normalization of variables to maintain consistent database in all observations and the comparable variables with each other. Several data preparation steps can be applied to ensure that the chosen explanatory variables have coherence with each other.

The discretization step is of fundamental importance, as while it is necessary to ensure that the phenomenon is correctly represented, classification methods often have a lot of numerical sensitivity, causing overfitting problems.

### 2.3. Modelling

During the modeling step, the techniques are identified and applied to preprocessed variables, and their results are examined in relation to the objective to be achieved. Several modeling techniques have been used, some of which require data to be in certain formats, which requires iterations with data preparation tasks (especially the discretization step).

In addition, configuration and calibration parameters may need to be specified for each modeling technique, and these are examined and optimized as much as possible during this step. Each technique has been implemented and parameterized independently to extract maximum efficacy and stability with different dataset subsamples.

### 2.4. Evaluation

The validation step involves the quantitative evaluation of the proposed models according to the degree in which they meet the objectives listed in the research hypotheses, and to what extent they are useful to correctly predict decision making. Hits/errors count was structured through the confusion matrix, so that the quality of the classification was earned using metric F1. Consistency test is an important tool for identifying overejustment problems or bias.

Comparison between methods transits between a quantitative and qualitative analysis, given that some methods can be poorly robust in some scenarios. Often a less specialized solution can be better generalizable.

## 3. Modelling and evaluation

### *3.1. Overview*

The stage of choice of indicators is of fundamental importance, however, there is no consensus on the literature on how to make this choice. Comparing several studies conducted in different countries, it is remarkable to realize that each economy is better represented by certain parameters and segments that have correlation with intrinsic characteristics of their economic activities. In other words, there is no specific methodology for this step.

The most common approach is to include the main indicators related to various segments of the economy, and later filter them through statistical correlation studies. Thus, the main indicators widely used in other studies were referenced, and an equivalent sought in the Brazilian economy.

In addition to the indicators without direct equivalent, some important indicators used during the research underwent methodology changes and their series were discontinued and restarted in other separate series, making a connection that maintained continuity. Therefore, the use of backup was important to maintain coherence at least in this first version of validation of the model.

### *3.2. Data preparation*

All variables chosen have monthly temporal granularity, and within the time interval chosen for the analysis, none of them presented missing data. However, each variable had a distinct unit of measure. The first transformation required to guarantee coherence and comparability between the variables was to consider the percentage variation between each monthly observation. As the purpose of the work is to predict the percentage variation of GDP, the percentage variation of each variable as a parameter for the classifier was adopted.

$$\Delta x_i = \frac{x_i - x_{i-1}}{x_{i-1}} \tag{1}$$

Some tests were performed considering the accumulated percentage variation, but the results found were not relevant. After normalizing the database, the pearson coefficients matrix (figure 2) was used for a linear correlation analysis with the GDP variable.

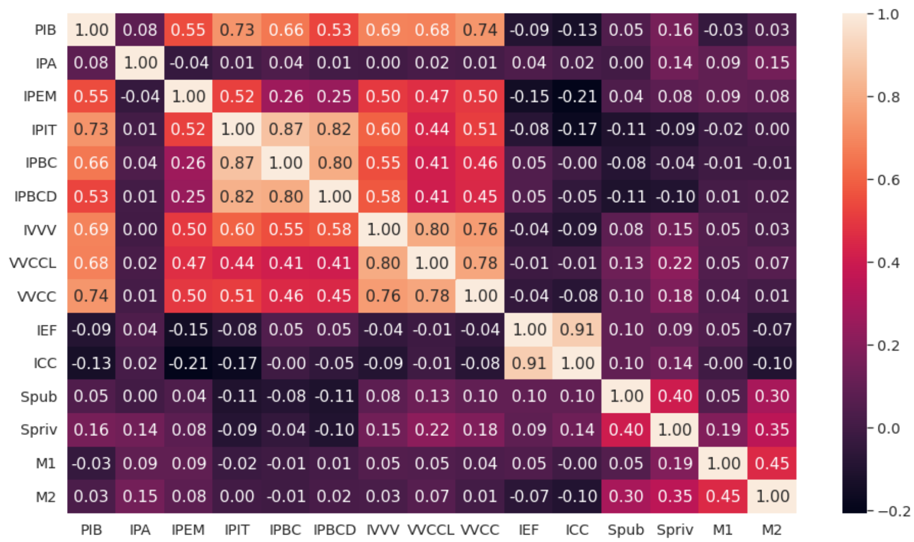


Figure 2: Correlations between the economic variables - complete dataset.

Although the correlation of several variables are apparently low, it must be remembered that the phenomena analyzed can be quite nonlinear. Seeking to consider these subtleties, a new database was proposed considering only the significant correlation variables, named as restricted dataset as shown as figure 3.

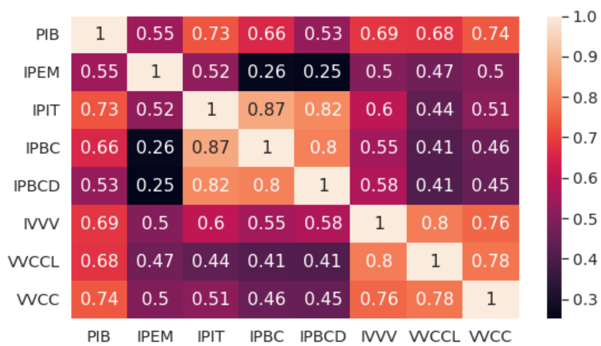


Figure 3: Correlations between the economic variables - restrict dataset.

To evaluate the series stationarity, KPSS (Kwiatkowski - Phillips - SCHMIDT - SHIN) and ADF (Augmed Dickey Fuller) tests were performed. Almost all original variables showed non-stationarity, however, all normalized variables showed stationarity in both tests (KPSS and ADF), which shows that the chosen variables have a stationarity in first difference.

### *3.3. Discretization*

The discretization step is a very important stage in the construction of the model, because influence on the correct visualization of the phenomenon. In this work several discretizations were proposed, seeking a sufficiently simple representation to avoid numerical problems, but capable of adequately representing decision -making conditions.

#### *3.3.1. Binary discretization*

Binary classification is a transformation that considers only the state of high and falling in each indicator, as shown in equation 2.

$$\begin{cases} x \mapsto 0, & \text{if } \Delta x < 0; \\ x \mapsto 1, & \text{if } \Delta x \geq 0. \end{cases} \tag{2}$$

Despite presenting excellent practical results, it is a very simplistic modeling that cannot separate regions of trend and stagnation, being little informative. Still, it was an important model to validate architecture and test the robustness of some techniques used.

#### 3.3.2. 3-categories discretization

Discretization with 3 categories separates significant high and falling movements with a stagnation movement, where the variation is less than a standard deviation in relation to the average, as shown in equation 3.

$$\begin{cases} x \mapsto -1, & \text{if } \Delta x \leq \mu_x - \sigma_x; \\ x \mapsto 0, & \text{if } \mu_x - \sigma_x < \Delta x < \mu_x + \sigma_x; \\ x \mapsto 1, & \text{if } \Delta x \geq \mu_x + \sigma_x. \end{cases} \tag{3}$$

This approach was very promising for a practical application, as it combines a good accuracy in the classification with effective categories in decision making.

#### 3.3.3. 5-categories discretization

The classification considering 5 categories is a refinement from 3 categories, as in addition to separating stagnation cases, it can also identify very rare cases. The first approach proposed for 5 categories is described in equation 4:

$$\begin{cases} x \mapsto -2, & \text{if } \Delta x \leq \mu_x - 2 \cdot \sigma_x; \\ x \mapsto -1, & \text{if } \mu_x - 2 \cdot \sigma_x < \Delta x \leq \mu_x - \sigma_x; \\ x \mapsto 0, & \text{if } \mu_x - \sigma_x < \Delta x < \mu_x + \sigma_x; \\ x \mapsto 1, & \text{if } \mu_x + \sigma_x \leq \Delta x < \mu_x + 2 \cdot \sigma_x; \\ x \mapsto 2, & \text{if } \Delta x \geq \mu_x + 2 \cdot \sigma_x. \end{cases} \tag{4}$$

Despite being a very interesting description for the phenomenon, as the extreme categories were very rare within the (relatively small) data set, training was not enough for methods to be able to satisfactorily identify these categories, causing great positive and negative fake and negative indexes, as well as convergence problems for some classifiers.

This discretization proposal Guaranteed 66% of data in the central category (stagnation) and only 2.5% of data reached the side categories. A new separation interval between the 5 categories was proposed, seeking to better balance these significantly rare events, as shown in equation 5:

$$\begin{cases} x \mapsto -2, & \text{if } \Delta x \leq \mu_x - 1,6745 \cdot \sigma_x; \\ x \mapsto -1, & \text{if } \mu_x - 1,6745 \cdot \sigma_x < \Delta x \leq \mu_x - 0,6745 \cdot \sigma_x; \\ x \mapsto 0, & \text{if } \mu_x - 0,6745 \cdot \sigma_x < \Delta x < \mu_x + 0,6745 \cdot \sigma_x; \\ x \mapsto 1, & \text{if } \mu_x + 0,6745 \cdot \sigma_x \leq \Delta x < \mu_x + 1,6745 \cdot \sigma_x; \\ x \mapsto 2, & \text{if } \Delta x \geq \mu_x + 1,6745 \cdot \sigma_x. \end{cases} \tag{5}$$

### *3.4. Classification*

#### *3.4.1. Cross-validation*

Before starting training, it is necessary to divide the training/test subsets, for application of cross validation that seeks to minimize the effects of overfitting and bias. As the variables are temporal series, it is important that the chronological order is not broken during cross validation. Therefore, the Time Series Split Cross Validation method was adopted, which initially considers small training/testing subsets at the beginning of the historical series, and follows in an iterative manner considering the previous round test subset as the current round training subset.

Several settings were tested for the size of the initial test subsets, but due to the large number of scenarios, in this work will be considered only the configuration that considers 30% for the subset of testing and validation.

#### *3.4.2. Classification mathods*

Several statistical and data mining techniques have been applied, which are available as part of the Python Scikit-Learn Library set. For all these techniques, the default settings proposed by the tool set were used, as the exhaustive optimization of the model's parameters and architectures was outside the scope of the article. However, seeking to solve some convergence problems, some of these techniques have been explicitly implemented to test some hyper parameters more attention. In detail, the techniques applied were: K-nearest neighbors (KNN), Gaussian naive bayes (NB), decision trees (DT), random forests (RF), logistics regression (LR), support vector machines (SVC), artificial neural networks (NN).

The K-nearest neighbors (KNN) is a method for classifying cases based on their similarity with other cases. The average or median target value of the nearest neighbors was used to obtain the expected value for the new case. Parameter K was automatically defined and the distance calculation metric was Euclidian. The naive bayes classifier (NB) is based on the application of the Bayes theorem with the "naive" assumption of conditional independence between each pair of characteristics given the value of the category variable. In this work was considered a Gaussian kernel whose coefficients were calculated through maximum likelihood. The decision tree (DT) is a map of the possible results of a series of related choices, while the random forest (RF) consists of building various decision trees considering different parts of the training set, seeking to reduce the estimator's variance and increasing accuracy. Logistic regression (LR) divides the dependent variable into two categorical levels, calculating the probability of output as a function of the input. In the proposed problem the input variables were also classified within these same cutting strips. The support vector classification (SVC) is a specific type of support vector machines (SVM) applied to classification problems and can be used to improve the predictive accuracy of a model without overfting training data and can be used to categorize a high space of dimensional appeal. It is noteworthy that this technique has brought excellent results, but it is quite sensitive to discretization, especially when trained with very rare and unbalanced categories. The artificial neural network (NN) can be used to automatically bring nonlinear relationships between inputs and the destination variable. The disadvantage of this flexibility is that the neural network is not easily interpretable. Specifically, the perceptron architecture of multi-charged was used, but the different discretizations brought a lot of variability in relation to hyperparimeters that were parameterized individually through pre-processing. Thus, the activation function, the solver, the number of neurons and hidden layers was parameterized on a case-by-case basis, to avoid problems of numerical convergence.

### 3.5. Target Variable Forecasting

To compare the various scenarios presented, a comparative analysis is proposed between all methods, in relation to the accuracy achieved in the training and testing steps. In addition, a more detailed analysis of the behavior of F1-score in relation to the different discretizations and the set of explanatory variables used in each scenario.

#### 3.5.1. Accuracy

To analyze accuracy, it is necessary to compare all the methods proposed with all discretization approaches.

Table 3: Comparison of accuracy in the training stage.

| Methods | 2CC | 2CR | 3CC | 3CR | 5CC | 5CR |
|---|---|---|---|---|---|---|
| kNN | 78,79% | 69,70% | 78,79% | 78,79% | 51,52% | 51,52% |
| NB | 75,76% | 72,73% | 84,85% | 78,79% | 57,58% | 57,58% |
| DT | 72,73% | 75,76% | 60,61% | 69,70% | 51,52% | 63,64% |
| RF | 72,73% | 72,73% | 78,79% | 69,70% | 54,55% | 54,55% |
| LR | 75,76% | 72,73% | 81,82% | 78,79% | 63,64% | 57,58% |
| SVC | 72,73% | 75,76% | 81,82% | 75,76% | 69,70% | 60,61% |
| NN | 72,73% | 75,76% | 81,82% | 75,76% | 63,64% | 57,58% |

Column labels refer to the discretization used and the amount of explanatory variables, for example:

- 2CC $\rightarrow$ 2 categories with the full base (2)
- 5CR $\rightarrow$ 5 categories with the restricted base (3)

Analyzing 3, we can see that there was considerable homogeneity in the measures, despite the diversity of methods and scenarios presented. Observing graphically, 4 illustrates the idea that there is a greater difficulty in classifying the most complex models, containing 5 categories.

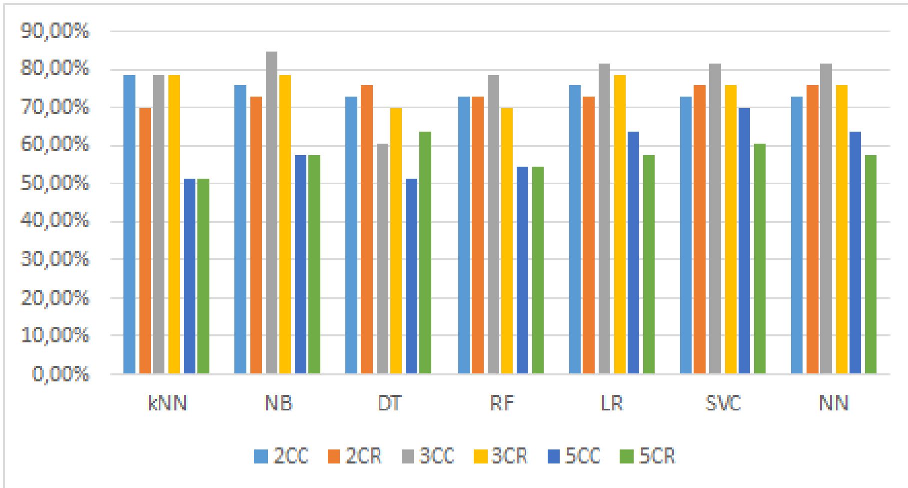


Figure 4: Comparison of accuracy in the training stage.

Comparing now the variability between methods considering the same discretization, 5 illustrates an interesting fact. For discretizations with 3 and 5 categories, reducing the number of explanatory variables brought lower variability of accuracy between methods.

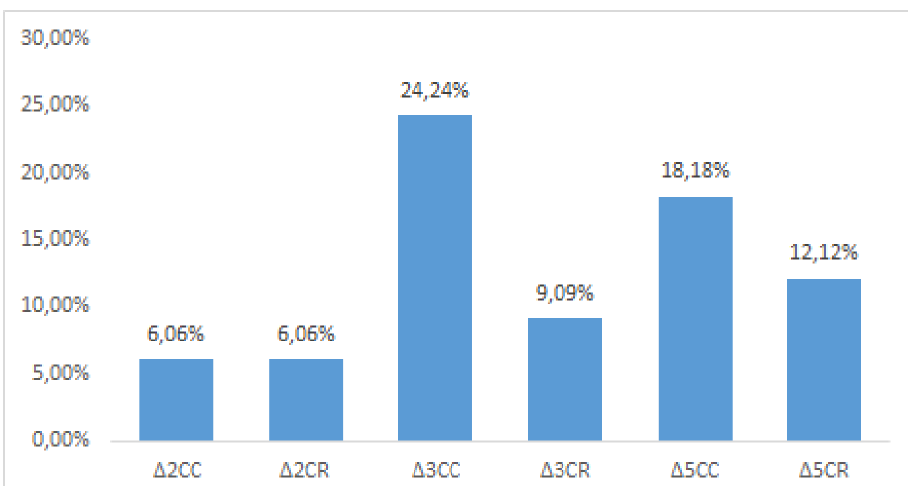


Figure 5: Variation of accuracy between methods in the training stage.

Repeating the approach to the test stage, we can evaluate the accuracy of all proposed scenarios.

Table 4: Comparison of accuracy in the test stage.

| Methods | 2CC | 2CR | 3CC | 3CR | 5CC | 5CR |
|---|---|---|---|---|---|---|
| kNN | 90,00% | 90,00% | 81,43% | 81,43% | 55,71% | 55,71% |
| NB | 87,14% | 87,14% | 78,57% | 80,00% | 50,00% | 51,43% |
| DT | 80,00% | 88,57% | 72,86% | 75,71% | 48,57% | 52,86% |
| RF | 88,57% | 90,00% | 82,86% | 75,71% | 65,71% | 60,00% |
| LR | 84,29% | 87,14% | 68,57% | 75,71% | 57,14% | 58,57% |
| SVC | 88,57% | 90,00% | 81,43% | 84,29% | 54,29% | 57,14% |
| NN | 85,71% | 85,71% | 84,29% | 82,86% | 58,57% | 60,00% |

As with the training stage, 4 shows that there was considerable homogeneity in the results obtained, despite the different discretizations. There was a slight improvement in the scenario considering binary classification, probably due to bias due to overly simple discretization for the analyzed phenomenon.

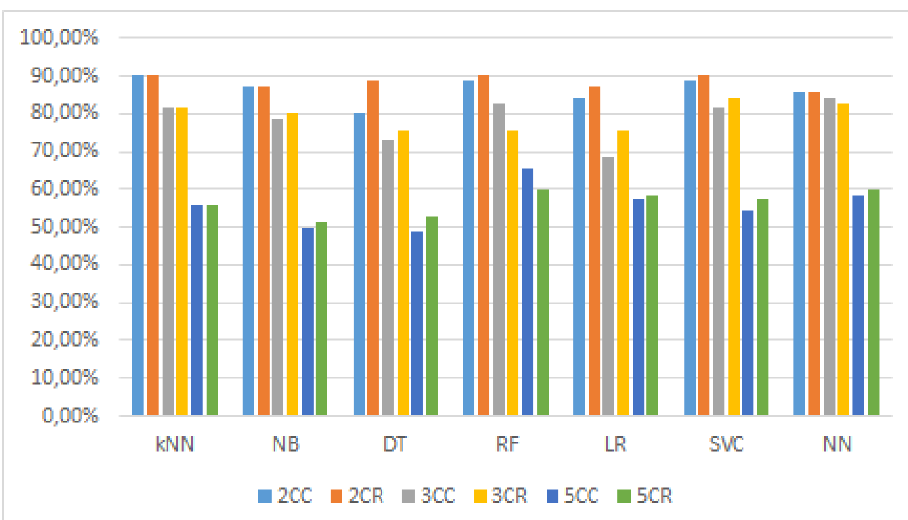


Figure 6: Comparison of accuracy in the training stage.

Again, observing accuracy related to the 5 categories, 6 reinforces the difficulty of forecasting in more complex models, also for the test step.

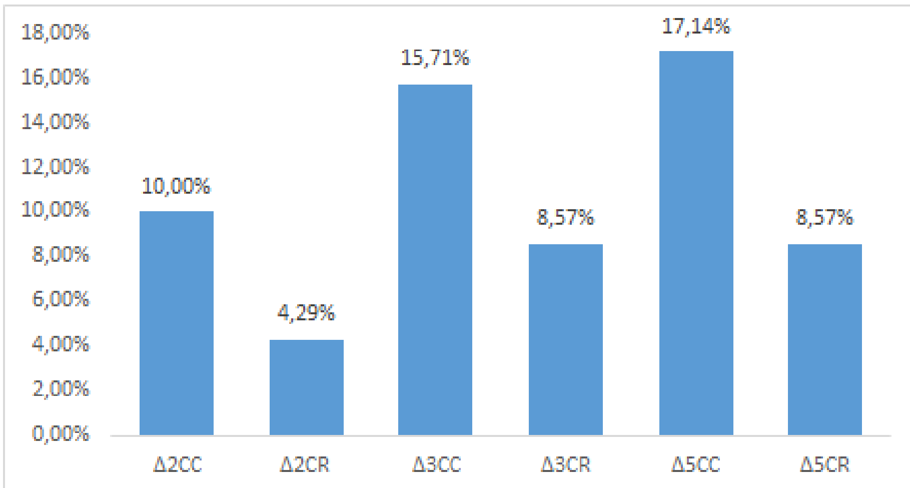


Figure 7: Variation of accuracy between methods in the training stage.

On the other hand, 7 makes it clear that reducing the number of explanatory variables has brought less variability between methods in the test step, regardless of the scenario.

*3.5.2. F-measure*

The quality of each method in relation to each scenario is relative, depending on the criterion. However, it is important to compare the effect that each scenario had on the performance of each percentage variation in the restricted base score F1 (some parameters) compared to the complete base (all parameters).As many scenarios were tested and in each of them there are many categories to evaluate, the number of results is significantly large. Thus, a qualitative analysis will be presented that demonstrates which the best scenario, which will also be discussed quantitatively.

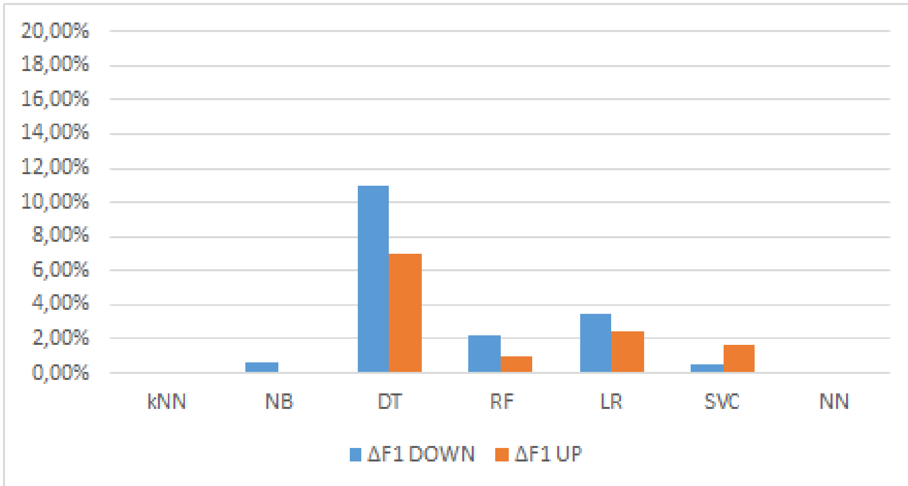


Figure 8: Score-f1 variation between methods with binary discretization.

Considering binary discretization, we can see from 8 that virtually all methods performed similarly when the restriction of explanatory variables applied. Although the decision tree has presented a greater improvement than other methods, 9% is not so significant. As the decision tree is a technique whose results can be interpreted (under certain circumstances), a more thorough analysis of this result is responsible.

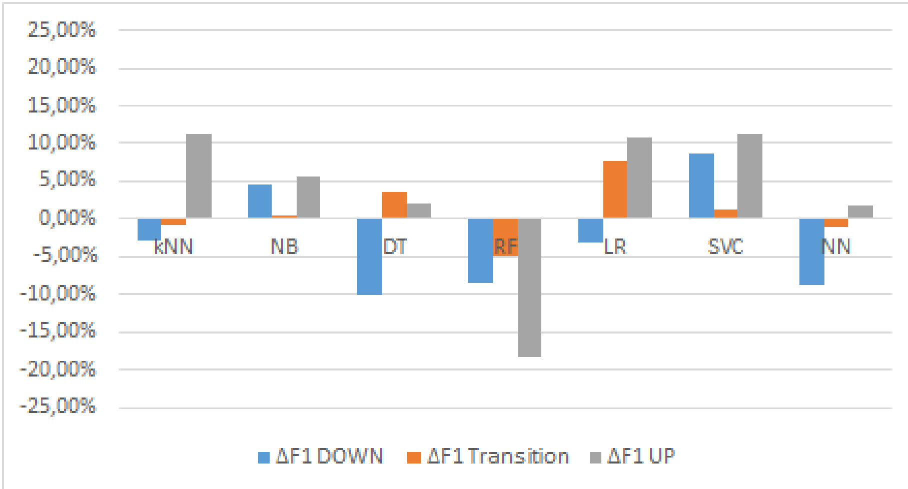


Figure 9: Score-f1 variation between methods with 3-categories discretization.

Evaluating the impact of the restriction of variables with the discretization of 3 categories, 9 shows that there was a non-negligible variation, but still within an average variation compatible with the previous case. As there were cases where there was improvement and others where it was worsening, it is not possible to say that the restriction is justified.

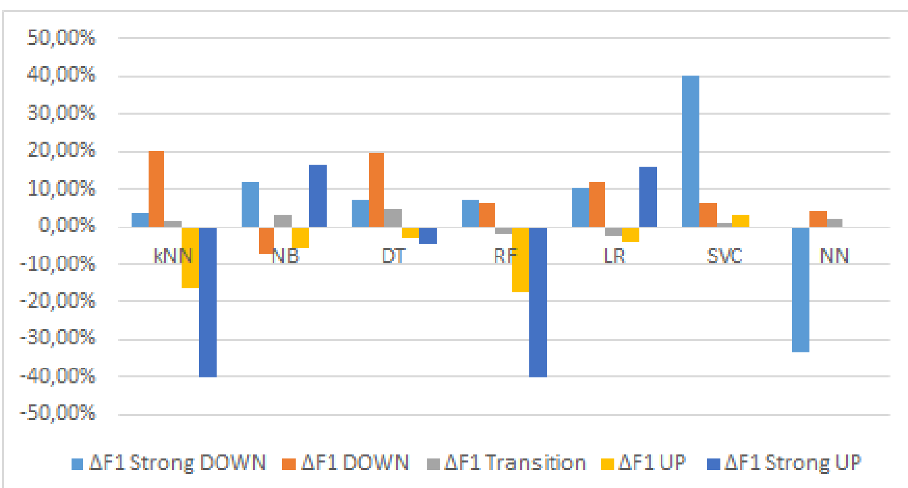


Figure 10: Score-f1 variation between methods with 5-categories discretization.

Finally, analyzing 10, we realize a relevant but not necessarily positive influence. On the one hand, the categories related to fall movements have a slight improvement (being very significant with respect to SVC), but on the other hand, the categories related to discharge movements have a significant worsening. However, it is important to emphasize that SVC and Neural Network did not converge to some categories, and the magnitude of worsening in some cases verges on 40%. This high variance suggests that limitation is much more linked to modeling errors than the predictive capacity of the methods.

Trying to summarize the total absolute gain achieved in each scenario when considering the restricted base, despite the high variance, 5 combines the absolute liquid gains of each category in relation to each scenario. As 7 methods are considered, the relative gain is precisely the average liquid gain between each of the 7 modeling.

Table 5: Comparison of score-f1 improvement when some parameters are removed.

| Categories | 2 categories | | 3 categories | | 5 categories | |
|---|---|---|---|---|---|---|
| | absolute | relative | absolute | relative | absolute | relative |
| -2 | | | | | 46,53% | 6,65% |
| -1 | | | -20,14% | -2,88% | 61,26% | 8,75% |
| 0 | 17,85% | 2,55% | 6,51% | 0,93% | 7,67% | 1,10% |
| 1 | 11,82% | 1,69% | 24,55% | 3,51% | -42,50% | -6,07% |
| 2 | | | | | -52,22% | -7,46% |
| Total | 29,67% | 4,24% | 10,92% | 1,56% | 20,74% | 2,96% |

Despite the significant improvement in the binary case, it has been discussed that it is a very simple approach that does not have a direct practical interest. On the other hand, the approach considering 5 categories, besides presenting a very chaotic performance, still had convergence problems, showing that the approach is not adequate.

For all that has been exposed, the scenario that considers 3 categories had an effective performance gain when the base was restricted to less significant variables. However, the relative improvement was very small, being not significant. Therefore, while removing them is a valid option, future studies may consider both cases in an attempt to verify in which cases they are significant.

Table 6: Testing sample results of the Brazilian GDP forecast step at the best model that considers 3 categories and the complete variable set

| Methods | Categories | Precision | Recall | F1 |
|---|---|---|---|---|
| | -1 | 80,00% | 36,36% | 50,00% |
| KNN | 0 | 79,66% | 97,92% | 87,85% |
| | 1 | 100,00% | 54,55% | 70,59% |
| | -1 | 62,50% | 45,45% | 52,63% |
| NB | 0 | 82,35% | 87,50% | 84,85% |
| | 1 | 72,73% | 72,73% | 72,73% |
| | -1 | 60,00% | 54,55% | 57,14% |
| DT | 0 | 82,22% | 77,08% | 79,57% |
| | 1 | 53,33% | 72,73% | 61,54% |
| | -1 | 71,43% | 45,45% | 55,56% |
| RF | 0 | 84,62% | 91,67% | 88,00% |
| | 1 | 81,82% | 81,82% | 81,82% |
| | -1 | 57,14% | 72,73% | 64,00% |
| LR | 0 | 88,24% | 62,50% | 73,17% |
| | 1 | 45,45% | 90,91% | 60,61% |
| | -1 | 80,00% | 36,36% | 50,00% |
| SVC | 0 | 79,66% | 97,92% | 87,85% |
| | 1 | 100,00% | 54,55% | 70,59% |
| | -1 | 83,33% | 45,45% | 58,82% |
| NN | 0 | 83,64% | 95,83% | 89,32% |
| | 1 | 88,89% | 72,73% | 80,00% |

Analyzing the values of 6, we realize that there is some variability in precision measurements, however, we will focus on the analysis only in metric F1. In any case, values are significant and can certainly be used as a tool in the study and characterization of the economic cycles phases.

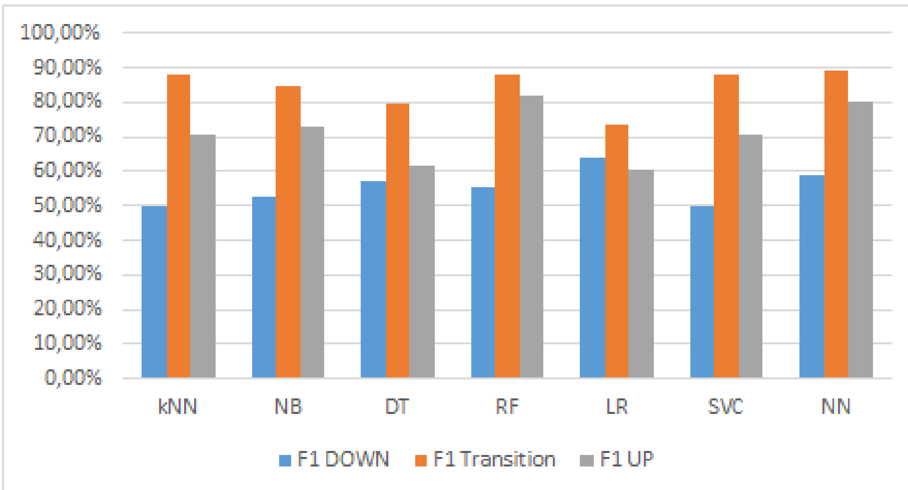


Figure 11: Score-F1 results of the Brazilian GDP forecast step at the best model that considers 3 categories and the complete variable set

We can see from 11 that all methods reach comparable accuracy, and virtually everyone is easier to predict the central category (stagnation), as well as a greater difficulty in predicting falls. This corroborates what is empirically observed, as low movements often happen faster and significantly, without leaving such clear signs. Despite this similarity with what is empirically observed, it is nonetheless a limitation of the model. Some methods (SVC and NN) have had problems of convergence and repeatability when the approach with 5 categories is used, as the Extreme Categories (Strong Up and Down) are quite unbalanced, causing model overfitting problems.

Another aspect that could influence overfitting or bias problems is the fact that the number of observations available is relatively small, as indicators have monthly temporal granularity. Therefore, several rounds of the same tests were performed with the same parameters to ensure the repeatability of the results.

Despite consistency in the observed results, it is always desirable to have more data available. Another possible approach to be investigated to improve the quality of the model is the inclusion of backward variables through autoregressive models.

## 4. Summary and conclusions

This article proposes a methodology to model and predict the value of Brazil's GDP, through the analysis of macroeconomic indicators related to different stages of the economic cycle. For this analysis, various statistical learning techniques are proposed, which seek different ways to aattack the nonlinearities of the model. This problem has a great interest from both an operational point of view, for monetary authorities and institutional market agents, as well as academic. Recent literature was very concerned with emphasizing the limitations of economic models, but made no effort to discuss aspects related to the interpretability of results. It is very important to offer a more systematic methodological support for implementing advanced techniques, as different work can be compared with each other.

The initial proposal to conduct a comparative study between various machine learning algorithms was exploited, but there was a greater focus on the data preparation stage, to facilitate the interpretation of the phenomenon. Because it is an unusual approach in the context of the Brazilian economy, besides the general understanding that could be applied in any market, it was necessary to find by attempt and error a configuration that adapted to the local context. Thus, the choice of variables as well as the steps of preparation and discretization were decisive not only to achieve good results, but also to interpret GDP behavior.

The proposed methodology was divided in a modular way to facilitate the understanding and implementation of continuous improvements. The variable categorization stage showed that simpler approaches (using fewer categories) often have better practical results, so it was necessary to find in the middle of simplicity and utility. The use of 3 categories can separate stages of stagnation and trend while ensuring good accuracy. Comparing the influence of low correlation variables, the model trained with fewer variables was slightly superior in general, but looking specifically for each category, it is not possible to establish a consensus since the forecasting methods vary significantly and become more unpredictable (smaller repeatability). Because of this, it is possible to state that despite the low linear correlation, these variables have significant importance in the model, especially when using techniques that can capture these nonlinear relationships.

Regarding the methods employed for forecast, all had similar behavior within the ideal scenario (3 categories with all variables). Some methods such as Support Vector Machines and Neural Networks had problems of convergence and overfitting when many categories were considered, especially because of the side categories (Strong Up and Down) were very unbalanced. Decision trees had a result had a slightly inferior accuracy, but can be preferable as the results can be interpreted more simply. Combined with random forests that always maintain a good result, this can be a set of efficient techniques that ensures good precision without losing the interpretative notion of the phenomenon. Data availability was not a limiting factor, but a larger amount of data available would probably bring better quality to the model. In addition, the thin adjustment of hyper parameters has brought little improvement in the results, showing that most of the methods employed are robust enough for this problem.

Finally, this study does not exhaust the possibilities on the subject, quite the contrary, opens many doors for new investigations. The proposed architecture can be used for other macroeconomic variables as well as other very different contexts of academic research and applications. In addition, made the appropriate considerations, the work could be extended to analyze a broader set of countries and considering other variables that connect these economies.

As well as the techniques proposed in this classification work did not have their possibilities exhausted, other machine learning techniques could be investigated, such as bagging and boosting techniques. In addition, other techniques may be better explored such as pruning random forests, or more sophisticated configurations of deep neural networks. It is noteworthy that the ideal configuration of the explanatory variables is still an open issue, enabling a deeper analysis in the face of the number of possibilities. Another approach to explore is the use of delay variables through self-regressive models, as it models the target variable not only as a function of the explanatory variable, but as part of an equation system simultaneously determined. In addition, techniques based on mobile averages that include components for seasonal trends are among the many future investigation possibilities.


### Acknowledgments

The authors would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Programa de Excelência Acadêmica (PROEX) - Brasil for Financial Support. We would also like to thank collaboration of Afonso de Campos Pinto and Elia Matsumoto (teachers and researchers at the Getulio Vargas Foundation School of Economics) and James Moraes de Almeida (professor and researcher at ILUM-CNPEM), for the review and suggestions for improvement in the Steps of the work.


### References


de A. Araújo, R., Oliveira, A.L., Meira, S., 2015. A hybrid model for high-frequency stock market forecasting. Expert Systems with Applications 42, 4081–4096. URL: https://www.sciencedirect.com/science/article/pii/S0957417415000093, doi:https://doi.org/10.1016/j.eswa.2015.01.004.

Alexander, G.J., Sharpe, W.F., Bailey, J.V., 2001. Fundamentals of investments. Pearson Educación.

Bhaumin, S., 2011 [Online]. Productivity and the economic cycle. BIS ECONOMICS PAPER NO. 12. URL: https://www.gov.uk/

government/uploads/system/uploads/attachment_data/file/32108/11-772-productivity-and-the-economic-cycle.pdf.

BROWN, S.J., DYBVIG, P.H., 1986. The empirical implications of the cox, ingersoll, ross theory of the term structure of interest rates. The Journal of Finance 41, 617–630. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1986.tb04523.x, doi:https://doi.org/10.1111/j.1540-6261.1986.tb04523.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1986.tb04523.x.

Burns, A.F., Mitchell, W.C., 1946. Measuring Business Cycles. National Bureau of Economic Research, Inc. URL: https://EconPapers.repec.org/RePEc:nbr:nberbk:burn46-1.

Cervelló-Royo, R., Guijarro, F., Michniuk, K., 2015. Stock market trading rule based on pattern recognition and technical analysis: Forecasting the djia index with intraday data. Expert Systems with Applications 42, 5963–5975. URL: https://www.sciencedirect.com/science/article/pii/S0957417415002110, doi:https://doi.org/10.1016/j.eswa.2015.03.017.

CFNAI, 2020. Chicago fed national activity index (cfnai) - federal reserve bank of chicago. URL: https://www.chicagofed.org/publications/cfnai/index.

Chang, P.C., Liu, C.H., Lin, J.L., Fan, C.Y., Ng, C.S., 2009. A neural network with a case based dynamic window for stock trading prediction. Expert Systems with Applications 36, 6889–6898. URL: https://www.sciencedirect.com/science/article/pii/S0957417408006209, doi:https://doi.org/10.1016/j.eswa.2008.08.077.

Chionis, D., Gogas, P., Pragidis, I., 2009. Predicting european union recessions in the euro era: The yield curve as a forecasting tool of economic activity. International Advances in Economic Research 16, 1–10. URL: https://doi.org/10.1007/s11294-009-9247-2, doi:10.1007/s11294-009-9247-2.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407. URL: http://www.jstor.org/stable/1911242.

COX, J.C., INGERSOLL JR., J.E., ROSS, S.A., 1981. A re-examination of traditional hypotheses about the term structure of interest rates. The Journal of Finance 36, 769–799. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1981.tb04884.x, doi:https://doi.org/10.1111/j.1540-6261.1981.tb04884.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1981.tb04884.x.

Diebold, F.X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337–364. URL: https://www.sciencedirect.com/science/article/pii/S0304407605000795, doi:https://doi.org/10.1016/j.jeconom.2005.03.005.

Enke, D., Thawornwong, S., 2005. The use of data mining and neural networks for forecasting stock market returns. Expert Systems with Applications 29, 927–940. URL: https://www.sciencedirect.com/science/article/pii/S0957417405001156, doi:https://doi.org/10.1016/j.eswa.2005.06.024.

ESTRELLA, A., HARDOUVELIS, G.A., 1991. The term structure as a predictor of real economic activity. The Journal of Finance 46, 555–576. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1991.tb02674.x, doi:https://doi.org/10.1111/j.1540-6261.1991.tb02674.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1991.tb02674.x.

Estrella, A., Mishkin, F.S., 1995. Predicting U.S. Recessions: Financial Variables as Leading Indicators. Working Paper 5379. National Bureau of Economic Research. URL: http://www.nber.org/papers/w5379, doi:10.3386/w5379.

Estrella, A., Mishkin, F.S., 1997. The predictive power of the term structure of interest rates in europe and the united states: Implications for the european central bank. European Economic Review 41, 1375–1401. URL: https://www.sciencedirect.com/science/article/pii/S0014292196000505, doi:https://doi.org/10.1016/S0014-2921(96)00050-5.

Estrella, A., Mishkin, F.S., 1998. Predicting U.S. Recessions: Financial Variables as Leading Indicators. The Review of Economics and Statistics 80, 45–61. URL: https://doi.org/10.1162/003465398557320, doi:10.1162/003465398557320, arXiv:https://direct.mit.edu/rest/article-pdf/80/1/45/1612423/003465398557320.pdf.

Gogas, P., Papadimitriou, T., Matthaiou, M., Chrysanthidou, E., 2014. Yield curve and recession forecasting in a machine learning framework. Computational Economics 45, 635–645. URL: https://doi.org/10.1007/s10614-014-9432-0, doi:10.1007/s10614-014-9432-0.

Guresen, E., Kayakutlu, G., Daim, T.U., 2011. Using artificial neural network models in stock market index prediction. Expert Systems with Applications 38, 10389–10397. URL: https://www.sciencedirect.com/science/article/pii/S0957417411002740, doi:https://doi.org/10.1016/j.eswa.2011.02.068.

Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. Econometrica 60, 77–105. URL: http://www.jstor.org/stable/2951677.

Hicks, J.R., et al., 1975. Value and capital: An inquiry into some fundamental principles of economic theory. OUP Catalogue .

HO, T.S.Y., LEE, S.B., 1986. Term structure movements and pricing interest rate contingent claims. The Journal of Finance 41, 1011–1029. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1986.tb02528.x, doi:https://doi.org/10.1111/j.1540-6261.1986.tb02528.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1986.tb02528.x.

Jacovides, A., 2008. Forecasting Interest Rates from the Term Structure: Support Vector Machines Vs Neural Networks. Master's thesis. University of Nottingham.

Ju, Y., Kim, C., Shim, J., 1997. Genetic-based fuzzy models: Interest rate forecasting problem. Computers & Industrial Engineering 33, 561–564. URL: https://www.sciencedirect.com/science/article/pii/S0360835297001939, doi:https://doi.org/10.1016/S0360-8352(97)00193-9. selected Papers from the Proceedings of 1996 ICC&IC.

Kamo, T., Dagli, C., 2009. Hybrid approach to the japanese candlestick method for financial forecasting. Expert Systems with Applications 36, 5023–5030. URL: https://www.sciencedirect.com/science/article/pii/S0957417408003321, doi:https://doi.org/10.1016/j.eswa.2008.06.050.

Kauppi, H., Saikkonen, P., 2008. Predicting u.s. recessions with dynamic binary response models. The Review of Economics and Statistics 90, 777–791. doi:10.1162/rest.90.4.777.

Kim, S.H., Noh, H.J., 1997. Predictability of interest rates using data mining tools: A comparative analysis of korea and the us. Expert Systems with Applications 13, 85–95. URL: https://www.sciencedirect.com/science/article/pii/S0957417497000109, doi:https://doi.org/10.1016/S0957-4174(97)00010-9.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curves. The Journal of Business 60, 473–489. URL: http://www.jstor.org/stable/2352957.

Oh, K.J., Han, I., 2000. Using change-point detection to support artificial neural networks for interest rates forecasting. Expert Systems with Applications 19, 105–115. URL: https://www.sciencedirect.com/science/article/pii/S0957417400000257, doi:https://doi.org/10.1016/S0957-4174(00)00025-7.

Rudebusch, G., Williams, J., 2009. Forecasting recessions: The puzzle of the enduring power of the yield curve. Journal of Business & Economic Statistics 27, 492–503. URL: https://EconPapers.repec.org/RePEc:bes:jnlbes:v:27:i:4:y:2009:p:492-503.

Stock, J.H., Watson, M.W., 1999. Forecasting inflation. Journal of Monetary Economics 44, 293–335. URL: https://www.sciencedirect.com/science/article/pii/S0304393299000276, doi:https://doi.org/10.1016/S0304-3932(99)00027-6.

Svalina, I., Galzina, V., Lujić, R., Šimunović, G., 2013. An adaptive network-based fuzzy inference system (anfis) for the forecasting: The case of close price indices. Expert Systems with Applications 40, 6055–6063. URL: https://www.sciencedirect.com/science/article/pii/S0957417413003151, doi:https://doi.org/10.1016/j.eswa.2013.05.029.

Svensson, L.E.O., 1994. Estimating and interpreting forward interest rates: Sweden 1992-1994. IMF Working Papers 94, 1. URL: https://doi.org/10.5089/9781451853759.001, doi:10.5089/9781451853759.001.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188. URL: https://www.sciencedirect.com/science/article/pii/0304405X77900162, doi:https://doi.org/10.1016/0304-405X(77)90016-2.

Vela, D., 2013. Forecasting Latin-American yield curves: An artificial neural network approach. BORRADORES DE ECONOMIA 010502. BANCO DE LA REPÚBLICA. URL: https://ideas.repec.org/p/col/000094/010502.html.

Zimmermann, H., Tietz, C., Grothmann, R., 2002. Yield curve forecasting by error correction neural networks and partial learning, in: Verleysen, M. (Ed.), ESANN 2002, 10th Eurorean Symposium on Artificial Neural Networks, Bruges, Belgium, April 24-26, 2002, Proceedings, pp. 407–412.