The effect of discretization on classification: a comparative study of machine learning methods applied to unusual discretization intervals for the characterization and prediction of economic variables

Eduardo Palhares Júnior[1], Rian Penachi[2], Adriano Honorato de Souza[3], Nivaldo Rodrigues e Silva[1], Wenndisson da Silva Souza[1], Edgard Gonçalves Cardoso[4]

[1] Federal Institute of Amazonas (Industrial District), Manaus 69075-351, Brazil
[2] University Center Fundação Santo André, Santo André 09060-650, Brazil
[3] Federal Institute of Amazonas, Itacoatiara 69101-030, Brazil
[4] Federal University of ABC, Santo André 09280-560, Brazil
eduardo.palharesjr@ifam.edu.br, rian.penachi@fsa.br , adriano.honorato@ifam.edu.br, nivaldo@ifam.edu.br, wenndisson.souza@ifam.edu.br, edgard.cardoso@ufabc.edu.br

Resumo

This work explores the classification of phases of the Brazilian economic cycle through a comparative study of different machine learning techniques. Models based on macroeconomic indicators were evaluated, focusing on the analysis and prediction of critical points in the cycle, such as the beginning of recessions and recoveries. An innovative approach was adopted when investigating unusual discretization ranges, aiming to reduce the impacts of outliers and problems of overfitting and bias in the data. The study highlights the importance of the discretization stage, especially in scenarios influenced by exceptional events, such as the COVID-19 pandemic, and analyzes the challenges in interpreting more abstract methods that, although effective, limit understanding of the underlying causes of models' behavior.

Palavras-chave: *Economic Cycle, Gaussian Distribution, Multiclass-discretization, Classification, Machine Learning.*

1. Introduction

Forecasting economic cycles remains a persistent challenge in both macroeconomics and applied machine learning. Accurate classification of cyclical phases such as expansion, stagnation, and recession plays a central role in economic planning, investment strategies, and public policy. However, the inherent non-linearity of macroeconomic indicators often limits the effectiveness of traditional models, particularly during transitional or crisis periods (Hamilton, 1989; Stock & Watson, 1999).

Recent advances in time series classification using supervised learning have shown promise in addressing this complexity (Bagnall et al., 2017). In Brazil, previous work by (Palhares Junior et al., 2024) proposed a framework combining macroeconomic indicators from the Central Bank's API with machine learning models to predict GDP dynamics through discretized class labels. Their findings highlighted the influence of discretization schemes on model performance, particularly in scenarios of heightened volatility and structural change.

During the presentation of this earlier study at an international conference, it was observed that the models systematically underperformed when forecasting sharp economic downturns—particularly the COVID-19 contraction. This behavior reflects a well-documented difficulty in modeling discontinuous or exogenous shocks, a topic broadly explored in financial risk literature (Cont, 2001; Taleb, 2007).

Motivated by these empirical asymmetries—and guided by feedback from economic theorists who questioned the assumption of non-symmetric cycle transitions—this paper introduces two major methodological advances. First, the temporal horizon of the dataset was extended to 2024, allowing the pandemic period to be included in model training. Second, new discretization strategies were explored, including a symmetric standard deviation-based approach and a hybrid percentile-based scheme, designed to better represent central and tail regions of the target variable.

These contributions originated from two final projects developed within a postgraduate program in machine learning, and were later consolidated into a single comparative study. The resulting analysis evaluates how alternative discretization frameworks affect the performance of multivariate time series classifiers trained to detect economic cycle phases. By focusing on the discretization stage - often overlooked in economic forecasting models - this paper contributes to a better understanding of how categorical representations can shape the predictive power of machine learning techniques in macroeconomic contexts.

2. Metodology overview

This study investigates the impact of class-label discretization strategies on the classification of Brazilian GDP cycles. The proposed pipeline builds upon the framework presented by (Palhares Junior et al., 2024), extending the temporal scope to include recent economic shocks and introducing alternative methods for labeling the target variable. The process is structured into four main stages: data acquisition and transformation, category modeling, model training, and performance evaluation.

Macroeconomic indicators were retrieved from the Central Bank of Brazil's API and represent a wide array of sectors, including industrial production, consumer confidence, vehicle sales, monetary aggregates, and credit balances from public and private financial institutions. The variables were converted into monthly variation rates to standardize the time series and reduce autocorrelation effects, as recommended in the time series forecasting literature (Hyndman & Athanasopoulos, 2018).

The target variable (monthly GDP growth) was discretized using three strategies: (i) a baseline scheme based on symmetrical intervals around the mean defined by standard deviation thresholds, resulting in three classes; (ii) refined symmetric intervals using alternative standard deviation multiples, and (iii) hybrid schemes defined by central and tail percentiles — both producing five-class configurations. The motivation behind using non-standard discretization methods stems from the literature on discretization for classification, which highlights the benefits of tailoring class boundaries to the empirical distribution of the data or the classification objective (Dougherty et al., 1995; Liu et al., 2002).

Supervised machine learning algorithms were applied to capture the relationship between the macroeconomic indicators and the discretized GDP classes. Performance evaluation was conducted using overall accuracy and macro-averaged F1-score, metrics widely recommended for multiclass and imbalanced classification problems (Provost & Domingos, 2003; Saito & Rehmsmeier, 2015).

2.1. Analysis and datapreparation

The dataset comprises monthly macroeconomic indicators provided by the Central Bank of Brazil, covering the period from January 2002 to May 2024. To ensure comparability across variables, all time series were transformed into monthly percentage changes, a common approach to capture short-term dynamics and mitigate issues related to non-stationarity and autocorrelation (Hyndman & Athanasopoulos, 2018; Palhares Junior et al., 2024).

Table 1: Description of 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)

The indicators listed in **Table 1** were selected to represent a broad set of economic dimensions, including production, consumption, credit, and expectations. This diverse composition aims to capture both real activity and forward-looking signals relevant to GDP behavior. The preprocessing step includes standardization, missing value handling, and feature transformation, ensuring that the data fed into the models is consistent and suitable for learning tasks.

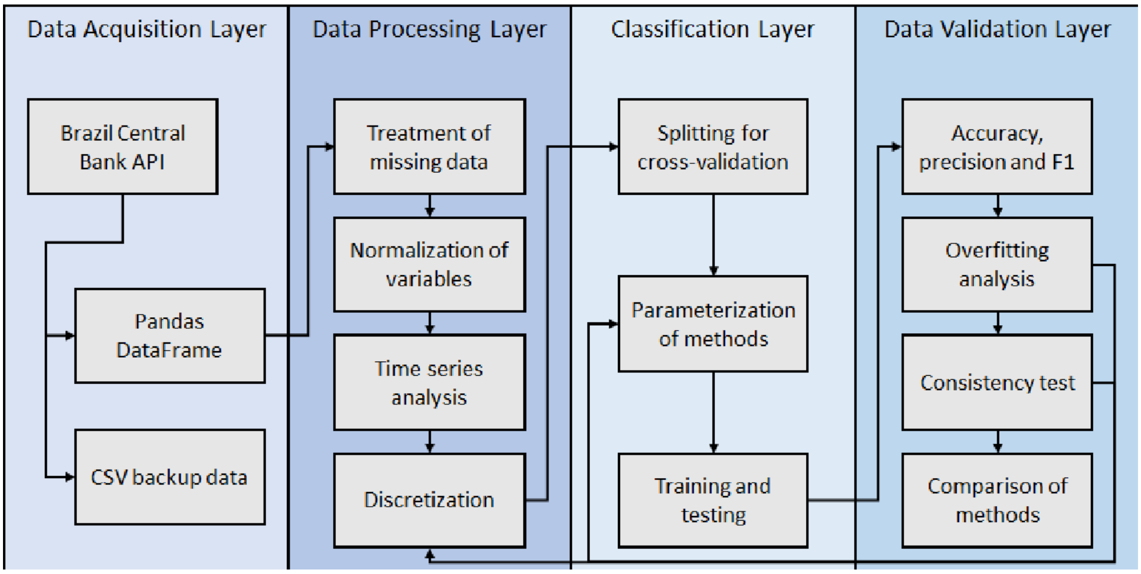


Figure 1: Methodological workflow for GDP cycle classification.

The proposed methodological workflow, summarized in **Figure 1:** Methodological workflow for GDP cycle classification., shows the data flow from raw macroeconomic series to final prediction outputs. In the modeling stage, various machine learning algorithms are trained using different discretization schemes applied to the target variable, enabling comparative performance evaluation.

2.2. Categories modeling

The forecasting task is formulated as a supervised classification problem, where the target variable corresponds to the monthly GDP variation. Since GDP is originally a continuous time series, different discretization strategies were applied to convert it into categorical labels representing economic cycle phases. This transformation enables the use of classification algorithms and allows performance comparisons across different class structures.

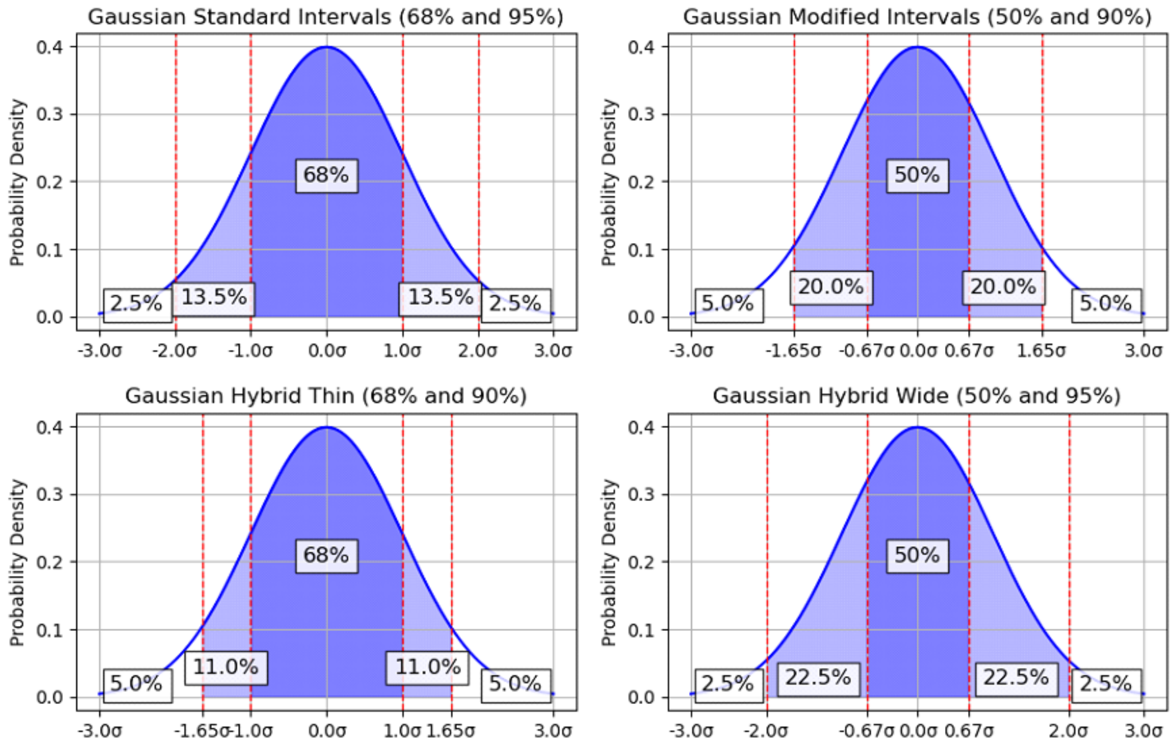


Figure 2: Discretization of a Gaussian distribution into five classes using varying thresholds.

To explore the impact of class structure on model performance, six discretization schemes were tested, including both three-class and five-class configurations. All strategies were based on symmetric intervals derived from the standard normal distribution, with variations in the threshold positions to alter the concentration of probability mass across classes. Although both three- and five-class setups were evaluated, only the five-class versions are graphically illustrated, since the three-class variants correspond to subcases obtained by removing the extreme categories.

These visual representations of the standard normal distribution clarify how each discretization strategy segments the probability space. As shown in **Figure 2**, the graphs illustrate how adjusting threshold positions influences the distribution of probability mass among classes, offering an intuitive understanding of the schemes evaluated in this study.

2.3. Forecast and validation

The experiments followed a temporal split: data from January 2002 to June 2014 was used for training, July 2014 to July 2017 for validation, and August 2017 to April 2024 for testing. This configuration ensures that the COVID-19 crisis, along with other recent shocks, is evaluated exclusively in the test set. Although the model had no prior exposure to pandemic data, the training period included milder disruptions such as the 2015–2016 impeachment proceedings and early signs of political and financial instability. These events may have contributed to the model’s robustness when facing more severe structural breaks.

All models were implemented in Python using the scikit-learn library. Standard hyperparameters were adopted, as prior experiments indicated negligible improvements from tuning in this context. Each model was trained using macroeconomic indicators as input and the discretized GDP variation as the target variable.

Performance was assessed using two complementary metrics: overall accuracy and macro-averaged F1-score. While accuracy provides a general view of correctness, the macro F1-score gives equal weight to each class, making it more informative under class imbalance - a recurrent feature in economic cycle classification (Provost & Domingos, 2003; Saito & Rehmsmeier, 2015).

3. Modelling and evaluation

This section presents the results obtained from the different class modeling strategies applied to the GDP classification task. Three discretization schemes (standard, modified, and hybrid) were evaluated under two temporal scopes: the original dataset (up to 2019) and the extended dataset (up to 2024), which incorporates the COVID-19 period.

The primary goal of the evaluation was to assess whether the choice of class thresholds has a measurable impact on predictive performance. In addition, we aimed to investigate whether including the pandemic period in the training set improves the model’s ability to detect abrupt economic downturns, a limitation observed in previous studies (Palhares Junior et al., 2024).

The results are presented in terms of overall accuracy and macro-averaged F1-score, both calculated on holdout test sets following the expanding window validation described in Section 2.3. To facilitate comparison, we grouped the results into tables by model type and discretization strategy. Visualizations were included to highlight specific behaviors, such as improvements in minority class performance or persistent confusion in extreme classes.

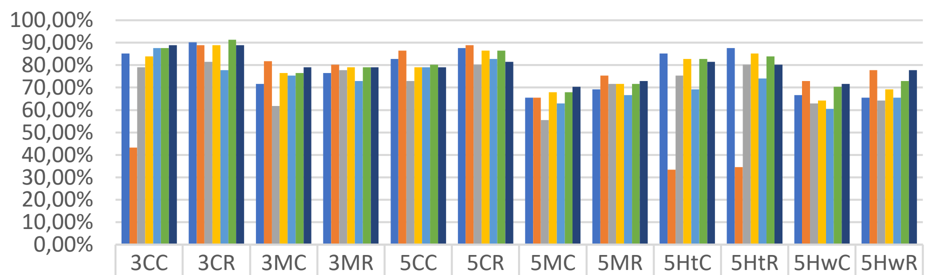
3.1. Accuracy

The classification accuracy observed during the test phase varied according to both the choice of discretization and the set of explanatory variables used. To provide a compact representation of all configurations, each column in the accuracy chart encodes three elements: the number of output classes (3 or 5), the discretization method, and whether the full set of macroeconomic indicators was used (“C”) or a restricted subset (“R”). The discretization codes include:

- C (Standard): based on symmetric thresholds around the mean, such as ±1σ for three classes or ±1σ and ±2σ for five classes;
- M (Modified): applied in both three and five-class configurations. In the three classes cases, it uses ±0.674σ (capturing 50% centrally). For five classes, it adopts ±0.674σ and ±1.645σ to represent 50% centrally, 40% intermediate (20% per side), and 10% in the tails;
- Ht (Hybrid Thin): uses ±1σ (68%) for the center and ±1.645σ (90%) for the outer limits, resulting in tighter intermediate zones;
- Hw (Hybrid Wide): combines ±0.674σ (50%) at the center and ±2σ (95%) at the extremes, emphasizing wider intermediate regions.

For example, the label “5HwR” denote to a five-class hybrid-wide discretization applied with a restricted variable set, while the rows indicate the classifiers used.

The results are summarized in **Figure 3**, which presents the test accuracy achieved by each model under every tested configuration. Among the classifiers, Random Forest and Gradient Boosting consistently yielded the highest accuracy across most discretization types. Simpler models such as Decision Trees and k-NN showed more variable performance, especially in five-class setups.



	3CC	3CR	3MC	3MR	5CC	5CR	5MC	5MR	5HtC	5HtR	5HwC	5HwR
kNN	85,19	90,12	71,60	76,54	82,72	87,65	65,43	69,14	85,19	87,65	66,67	65,43
NB	43,21	88,89	81,82	80,25	86,42	88,89	65,43	75,31	33,33	34,57	72,84	77,78
DT	79,01	81,48	61,73	77,78	72,84	80,25	55,56	71,60	75,31	80,25	62,96	64,20
RF	83,95	88,89	76,54	79,01	79,01	86,42	67,90	71,60	82,72	85,19	64,20	69,14
LR	87,65	77,78	75,31	72,84	79,01	82,72	62,96	66,67	69,14	74,07	60,49	65,43
SVC	87,65	91,36	76,54	79,01	80,25	86,42	67,90	71,60	82,72	83,95	70,37	72,84
NN	88,89	88,89	79,01	79,01	79,01	81,48	70,37	72,84	81,48	80,25	71,60	77,78

Figure 3: Test accuracy across models and discretization schemes.

In terms of discretization, the standard three-class (3C) configuration achieved the highest accuracy overall, as expected from its lower complexity and more balanced class distribution. As the number of classes increases, accuracy tends to decrease, reflecting the the added difficulty of the task. However, within the five-class configurations, modified and hybrid strategies such as 5M and 5Ht often outperformed the standard 5C configuration, suggesting that redistributing probability mass more evenly can enhance model learning.

The use of a restricted variable set ("R" configurations) generally produced comparable or slightly better accuracy than the full set ("C"), indicating that removing weakly correlated indicators helped reduce noise without sacrificing predictive power.

3.2. F-Score

While overall accuracy offers a general view of model correctness, it may obscure important disparities in class-wise performance — especially in imbalanced classification settings. To address this, we analyze both the macro-averaged F1-score and the class-wise F1-scores under each discretization strategy. The macro-F1 is particularly appropriate in economic contexts, as it assigns equal importance to each class regardless of prevalence (Saito & Rehmsmeier, 2015).

Given the large number of configurations and models, displaying full sets of F1-score results would be redundant and uninformative. Instead, we prioritize a comparative analysis of the effect of using a restricted versus complete set of explanatory variables. **Table 2** summarizes the relative improvement (%) in class-wise F1-scores when switching from the complete ("C") to the restricted ("R") base. Each row corresponds to a class, and each column represents a discretization scheme.

Table 2: Relative improvement (%) in class-wise F1-score when applying variable restriction.

Categories	3 standard	3 modified	5 standard	5 hybrid 50/90	5 hybrid 68/90	5 hybrid 50/95
-2			0,00%	21,28%	-66,68%	0,00%
-1	81,11%	21,82%	116,41%	28,30%	48,20%	48,22%
0	48,13%	15,42%	18,35%	44,13%	2,21%	31,15%
1	46,32%	45,87%	104,75%	60,95%	120,73%	33,44%
2			52,67%	-12,46%	-108,70%	-100,31%
Total	175,56%	83,12%	292,18%	142,21%	-4,24%	12,50%

The results confirm that, for three-class configurations, variable restriction consistently improves model performance across all classes. This validates the use of restricted models as reference in this context. In five-class scenarios, the impact of restriction varies: the standard (5C) and modified (5M) discretizations benefit from reduced dimensionality, while hybrid schemes such as 5Ht (68/90) and 5Hw (50/95) exhibit substantial degradation in performance - especially in rare classes. These findings indicate that more complex class boundaries may demand richer feature sets to avoid excessive sparsity in the tails.

To illustrate the best-case scenario, **Figure 4** shows the class-wise F1-scores for the 3CR configuration (three-class standard discretization with restricted variable base). This setup demonstrated consistently strong performance across all three classes, particularly in class 0 and class 1. The combination of standard symmetric thresholds (±1σ) and variable restriction provided a robust and balanced model structure, yielding macro-F1 scores comparable to the top-performing models. These results suggest that conventional discretization can be highly effective when paired with a carefully selected feature set.

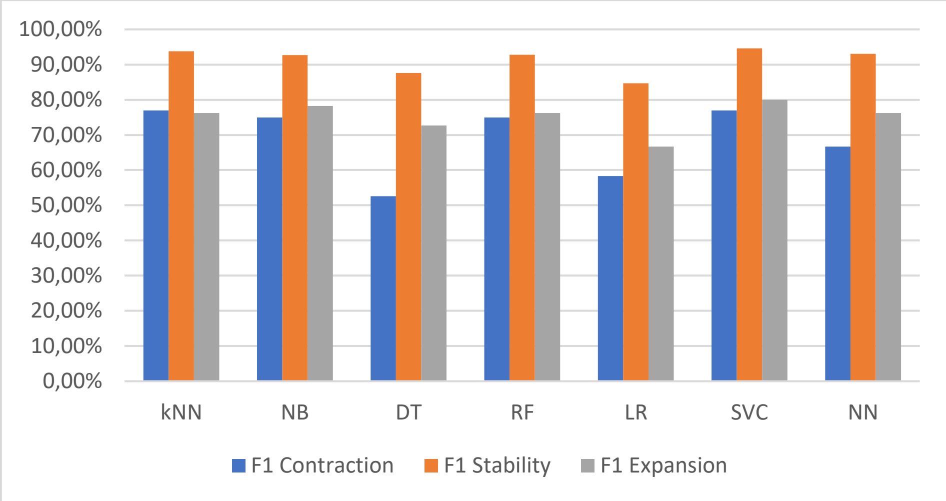


Figure 4: Class-wise F1-scores for the 3-class standard discretization using the restricted variable base (3CR).

In contrast, **Figure 5** presents the class-wise F1-scores for the 5HtR configuration (five-class hybrid-thin 68/90 discretization with restricted variables), which yielded the worst overall results. Here, the outer classes (-2 and 2) are poorly identified, with F1-scores near zero for models such as SVC and Neural Networks. The threshold ranges for both intermediate and extreme classes appear too restrictive given the limited feature set, leading to model underfitting and classification instability.

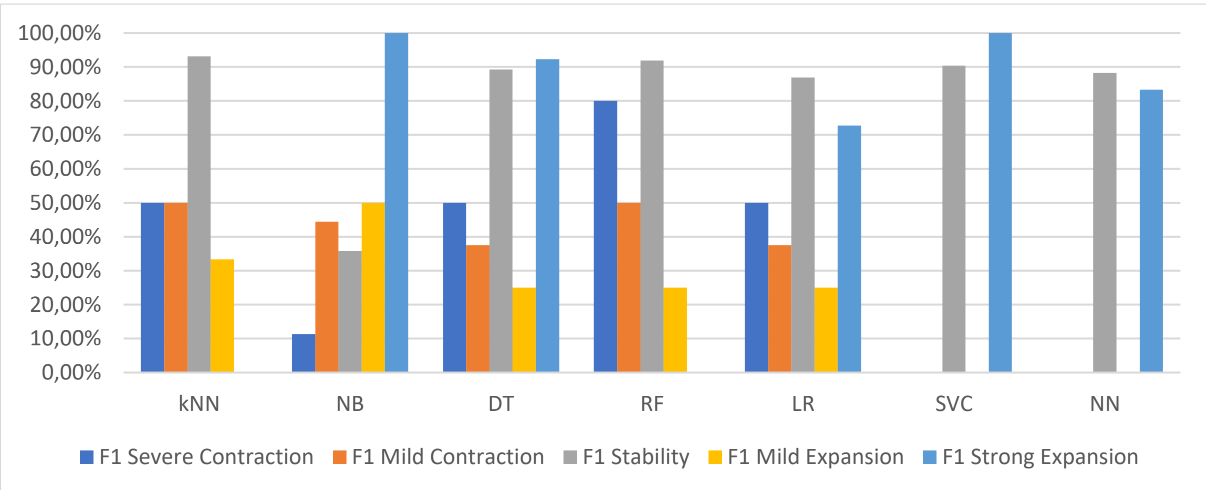


Figure 5: Class-wise F1-scores for the 5-class hybrid-thin discretization (68/90) using the restricted variable base (5HtR). Performance is especially poor in extreme classes.

These contrasting examples highlight the importance of aligning discretization complexity with the richness of the feature space. While variable restriction enhances robustness in simpler scenarios, it can hinder performance when class structures require more nuanced distinctions. As such, feature selection should be guided by both empirical performance and the underlying data distribution.

4. Summary and conclusions

This study presented a comparative analysis of discretization strategies for classifying monthly GDP variation using supervised machine learning. The investigation was motivated by earlier work on economic cycle prediction, which was expanded and refined through two successive graduate-level projects. The consolidated approach enabled both methodological generalization and empirical depth, incorporating findings from prior experiments and key feedback received in academic forums.

The core research question was whether non-standard discretization schemes—designed to align more closely with the empirical distribution of GDP growth—could improve classification performance relative to conventional symmetric thresholds. Six strategies were tested, ranging from standard terciles to hybrid percentile-based thresholds, across multiple classifiers using both full and restricted sets of macroeconomic indicators.

Results showed that simpler setups, such as the three-class standard discretization with restricted variables (3CR), consistently delivered strong performance and model stability. This supports the notion that conventional thresholds remain effective when combined with targeted feature selection. In contrast, more complex five-class discretizations—especially hybrids like 5Ht (68/90)—performed poorly when paired with limited feature sets. These results suggest that intricate class boundaries demand richer inputs to avoid sparsity and instability, particularly in underrepresented regions.

These findings also help clarify ambiguities from earlier experiments. Prior results indicated an asymmetry in model performance for downturns versus upswings—an interpretation initially at odds with economic theory. To investigate further, the study was extended to a more recent time window, incorporating events such as the 2016 political crisis, the 2017 recession, and the 2018 truck drivers' strike. Although the COVID-19 pandemic remained outside the training window, its presence in the test set—along with prior exposure to other shocks—likely improved the model's ability to generalize to extreme scenarios.

This raises a broader methodological debate: whether rare disruptive events should be excluded to reduce bias or included to enhance resilience. While some recommend removing outliers to avoid overfitting, others—like the present study—argue that exposure to structural breaks is essential in economic modeling, where shocks are not anomalies but recurring features of the system.

Overall, this study reinforces the need to match model complexity with data quality. While exotic discretization methods may offer theoretical advantages, their practical gains depend on the availability of meaningful, structured features. Future research could explore dynamic or data-driven thresholding methods, as well as the impact of explicitly excluding structural outliers like COVID-19 to compare generalization outcomes. A promising path involves integrating this framework with NLP tools and large language models (LLMs), enabling hybrid systems that combine structured indicators with unstructured data. Techniques such as retrieval-augmented generation (RAG) and fine-tuning may further improve forecasts by embedding economic reasoning into the predictive process.

Acknowledgments

The authors would like to thank SAMSUNG ELETRÔNICA DA AMAZÔNIA LTDA. through the Projeto Aranouá and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Programa de Excelência Acadêmica (PROEX) - Brasil for Financial Support. The present work is the result of the Research and Development (R&D) project 001/2021, signed with Instituto Federal do Amazonas and FAEPI, Brazil, which has funding from Samsung.

References

Bagnall, A., Lines, J., Bostrom, A., Large, J., & Keogh, E. (2017). The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, *31*(3), 606–660. https://doi.org/10.1007/s10618-016-0483-9

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, *1*(2), 223–236. https://doi.org/10.1080/713665670

Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. *Proceedings of the 12th International Conference on Machine Learning*, 194–202.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*(2), 357–384. https://doi.org/10.2307/1912559

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2° ed). OTexts. https://otexts.com/fpp2/

Liu, H., Hussain, F., Tan, C., & Dash, M. (2002). Discretization: An enabling technique. *Data Mining and Knowledge Discovery*, *6*(4), 393–423. https://doi.org/10.1023/A:1016304305535

Palhares Junior, E., Araujo, A. M. T. de, Souza, A. H. de, Silva, N. G. da, & Souza, W. da S. (2024). Ensemble of machine learning applied to economic cycles analysis: A comparative study using antecedent macroeconomic indicators for Brazilian GDP prediction classification. *Revista Brasileira de Planejamento e Desenvolvimento*.

Provost, F., & Domingos, P. (2003). Tree induction for probability-based ranking. *Machine Learning*, *52*(3), 199–215. https://doi.org/10.1023/A:1024090213889

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Stock, J. H., & Watson, M. W. (1999). Business cycle fluctuations in US macroeconomic time series. Em *Handbook of Macroeconomics* (Vol. 1, p. 3–64). Elsevier.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*. Random House.