NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Comparative Analysis of GDP Forecasting using Ensemble Tree Regression Models:

Machine Learning vs. Econometric Models

Ricardo Paulo Barbosa de Carvalho Almeida Coelho

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Comparative Analysis of GDP Forecasting using Ensemble Tree Regression Models:
Machine Learning vs. Econometric Models

by

Ricardo Paulo Barbosa de Carvalho Almeida Coelho

Master Thesis presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Business Analytics.

Supervised by

Prof. Jorge Miguel Ventura Bravo, PhD

NOVA Information Management School &

Université Paris-Dauphine PSL

July, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, 07/07/2024

ABSTRACT

This thesis evaluates whether machine learning can have better results than well-established institutions in forecasting Portugal's GDP growth using ensemble tree regression models, with the OECD's economic outlook forecasts for Portugal serving as a benchmark based on data from 1962 to 2022 recovered from OECD and the Federal Reserve's economic data. The findings reveal that, in general, machine learning did not surpass the forecasts of the OECD. However, machine learning demonstrated the potential for better accuracy at many points, despite a higher propensity for larger errors compared to traditional methods. Among the ensemble tree regression models tested, the gradient boosting regressor consistently provided the best forecasts across all horizons, outperforming the random forest, extreme gradient boosting, and light gradient boosting machine models. The results also suggest that machine learning performs better with a larger volume of data than with higher dimensionality, even if some data points seem irrelevant to forecast future values. This thesis highlights the potential of machine learning in time series forecasting as a complementary tool to traditional methods, rather than a complete replacement.

KEYWORDS

Machine learning; Forecasting; GDP; Portugal; Gradient Boosting Regressor

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ADE	Arbitrated Dynamic Ensemble
AR	Auto-Regressive
ARIMA	Auto-Regressive Integrated Moving Averages
BMA	Bayesian Model Averaging
CPI	Consumer Price Index
CRISP-DM	Cross Industry Standard Process for Data Mining
GDP	Gross Domestic Product
IMA	Integrated Moving Average
LGBM	Light Gradient Boosting Machine
MIMO	Multi-Input Multi-Output
ML	Machine Learning
MLP	Multi-Layer Perceptron
MPE	Mean Percentage Error
NiGEM	National Institute Global Econometric
NN	Neural Network
OECD	Organization of Economic Cooperation and Development
R^2	R-squared
RF	Random Forest
Acronym RMSE	Root Mean Squared Error
Acronym SARIMA	Seasonal Autoregressive Integrated Moving Average
Acronym SVM	Support Vector Machine
Acronym US	United States
Acronym VAR	Vector Auto-Regressive
Acronym XGB	Extreme Gradient Boosting

1. INTRODUCTION

Predicting future trends is a vital part of decision-making that goes beyond just policymakers, affecting investors, managers, and everyday people. When forecasts are inaccurate or delayed, the consequences can be substantial, resulting in less-than-ideal decisions based on incorrect predictions (Jung et al., 2018). Therefore, the need for prompt and accurate forecasts is paramount, as they are instrumental in forming effective and well-informed decisions, ultimately benefiting our society.

Machine Learning (ML) is a valuable tool to improve the accuracy of forecasts produced by established institutions. While traditional forecasting techniques are widely used, they sometimes fall short due to inherent limitations, such as rigid assumptions. Recent research indicates that ML can address these limitations and potentially enhance forecasting precision, helping institutions produce more reliable predictions (Bolhuis & Rayner, 2020; Jung et al., 2018; Yoon, 2021).

However, ML, while being a powerful tool, should not be seen as a direct substitute for traditional methods. Its inherent nature, defined by algorithms, makes fully understanding the forecast's full scope and the relationships between variables and the target difficult, encapsulating it as a "black box". As a result, ML should be viewed as an additional instrument in the forecasting toolkit for policymakers (Haque, 2023). This thesis will primarily focus on the forecast output rather than extensively exploring the complex relationships between variables and the target.

This thesis will concentrate on a comparative examination of Portugal's Gross Domestic Product (GDP) growth forecasting using ML and Organization for Economic Cooperation and Development’s (OECD) forecast. Despite the existence of earlier research on a similar theme, our focus will be on filling a significant research gap related to the use of ensemble models based on regression trees. The central question driving this study is: "Can ML techniques, specifically ensembles of regression trees, arrive at better results than OECD’s forecast for the GDP growth of Portugal?" The primary goal is to forecast Portugal's GDP using ML with ensembles of regression trees and then compare these forecasts with the results produced by the OECD, which uses econometrics.

The methodology used to develop the models for this study adheres to the Cross Industry Standard Process for Data Mining (CRISP DM). Our strategy involves using an ensemble of ML algorithms, with a special focus on regression trees and their variations. These models are trained and evaluated using historical GDP data from 1962 to 2019 and a wide range of macroeconomic variables. To establish a meaningful comparison with the OECD's forecast, we use evaluation metrics such as Root Mean Squared Error (RMSE), Mean Percentage Error (MPE), and R-squared (R^2). Our choice of these methods and tools is supported by a comprehensive review of existing literature and practical implementations, ensuring a thorough and accurate evaluation of GDP forecasts for Portugal.

The choice of the data period from 1962 to 2019 to train and evaluate the models is motivated by significant global events that have had a profound impact on the world. Events like the Covid-19 pandemic, among others, have had a substantial effect on economic conditions, making the future highly unpredictable. The forecasts themselves focus on pre-pandemic quarters such as the last 2 quarters of 2019[1], but also the last two quarters of 2021, and the entire year of 2022. This approach bypasses the extreme outlier that was the year of 2020 in terms of GDP growth.

The expected results of this study aim to highlight the benefits of using ML methods for predicting macroeconomic indicators. Beyond this, our study aims to encourage wider adoption of sophisticated data-driven methodologies in economic forecasting within Portugal. By demonstrating the potential benefits of modern computational techniques, we aim to advance the field and contribute to its development.

The thesis is organized into six sections: literature review, methodology, data, machine learning techniques and model, results, and conclusions. The literature review explores previous research on forecasting with ML. The methodology section is divided into two subsections: the first details the methodology implemented for our forecast and its comparison to econometric methods, while the second explains the OECD's forecasting methodology. The data section describes the data used for our forecast, including its limitations and any transformations applied. The machine learning techniques and model section covers the techniques and methods used in constructing the forecast models. This includes handling outliers, feature selection, the functioning of the Gradient Boosting Regressor (GBR) algorithm, the rolling window technique, and the multiple steps ahead approach. The results section presents the findings of our forecast. Finally, the thesis concludes with the summary of the conclusions.

[1] The Covid-19 pandemic started on the 12th of December of 2019 in China (Prevention, 2019) but it only started influencing the world economy in 2020 (The World Bank, 2022).

2. LITERATURE REVIEW

Forecasting is a vital component in macroeconomics, offering critical insights for predicting the future state of the economy and its interplay with the global environment. This process is crucial for facilitating prompt responses and the creation of effective policy measures to uphold economic and financial stability, as well as aiding in the prevention of recessions (Jung et al., 2018). Moreover, forecasting holds intrinsic significance within macroeconomics, primarily due to its capacity to supply policymakers with timely and pertinent information. Even slight delays in forecasts, such as a one-month lag, can result in less-than-ideal decision-making and negative outcomes (Soybilgen & Yazgan, 2021; Yoon, 2021).

In this literature review, we will delve into the articles that discuss forecasting using ML relative to methods utilized by well-established international institutions, mainly econometrics, and their relevance in the sphere of macroeconomics. Our exploration will begin by detailing the existing methodologies and their associated limitations. Following this, we will examine ML models and their respective strengths. Finally, we will conduct a comparative analysis of the results derived from both approaches. In addition, we will clarify the reasoning behind our preference for ensemble regression tree models over other methods and conclude this section with a succinct summary of our findings.

Forecasting techniques derived from econometrics include a broad range of models, such as Auto-Regressive (AR), Vector Auto-Regressive (VAR), Auto-Regressive Integrated Moving Averages (ARIMA), Seasonal Autoregressive Integrated Moving Averages (SARIMA), and Integrated Moving Average (IMA), among others. Each of these models has its own limitations, including their focus on individual variables, dependence on linear assumptions, the requirement for data stationarity, potential neglect of key parameters, and their inability to capture nonlinear relationships between variables (Kaushik, 2020; Bolhuis & Rayner, 2020; Jung et al., 2018; Yoon, 2021).

These assumptions may seem increasingly unrealistic today, given the growing complexity and volume of data, as well as the rising uncertainties in macroeconomic factors, financial dynamics, and housing market bubbles. Economists are acknowledging the growing need for the integration of ML as a valuable tool (Jung et al., 2018).

In recent years, numerous articles have compared the results of traditional methods for forecasting methods used by reputable institutions with independently developed ML models. These comparisons consistently show that the ML methods are not only as effective as established techniques but also offer enhancements.

One of the models widely used in this context is the Neural Network (NN) and its variations. Maehashi & Shintani (2020) used NN to predict Japanese macroeconomic variables such as consumer price index and wholesale price index and noted superior predictive power in both medium and long-term forecasts. Haque (2023) also utilized NN to predict retail demand in

Turkey, leveraging a diverse dataset that includes unconventional macroeconomic variables to improve the accuracy of traditional forecasting methods.

Several factors contribute to the success of NN and its variants. NN excels at mapping dynamic non-stationary time-series due to its non-parametric nature, absence of rigid assumptions, tolerance for noise, adaptability, and its ability to capture randomness (Kaushik, 2020).

However, NN has its own shortcomings, such as a tendency to overfit, error backpropagation, and the need to determine optimal lag times. As a result, researchers explore alternative models to enhance forecasting accuracy. Kaushik (2020) also applied Support Vector Machine (SVM) to predict foreign exchange rates and, like NN, found that it outperforms traditional methods like VAR and ARIMA. SVM aims to minimize an upper bound of the generalization error rather than just the training error, although it may underperform with noisy and large datasets.

There are several ML models, each with its own set of strengths and weaknesses, excelling in specific domains and contexts. The key takeaway is that ML models outperform traditional methods by emphasizing out-of-sample performance and handling nonlinear interactions more effectively. This leads to superior forecasting of complex nonlinear relationships between variables while also mitigating the risk of over-extrapolating historical data (Bolhuis & Rayner, 2020). Consequently, as more data is provided to the model, the accuracy of forecasts surpasses that of traditional econometric models (Coulombe et al., 2020).

While it is true that the effectiveness of ML varies based on the data, context, and situation, with each method having its own advantages and disadvantages, more recent studies appear to show that ensemble tree regression models outperform the models described. As Maehashi & Shintani (2020) concluded in their forecast of Japan's macroeconomic variables, this approach considers not only NN but also incorporates Random Forest (RF), Bagging, and Boosting, consistently achieving better results compared to individual ML models. Similarly, Bolhuis & Rayner (2020) reached the same conclusion when forecasting Turkey's growth. They analysed RF, SVM, and GBT to compare the strengths of each model. While RF excelled in overall forecasting capability, it failed to capture significant growth fluctuations. SVM was better at capturing these situations but tended to overestimate them. GBT, on the other hand, also captured growth fluctuations but with less volatility than SVM. Their findings suggested that ensemble tree regression algorithms reduce forecast errors and are more adept at identifying volatility compared to their counterparts.

Bravo and El Mekkaoui (2022) apply a flexible Bayesian model averaging (BMA) approach to Consumer Price Index (CPI) inflation forecasting to mitigate conceptual uncertainty and improve the short-term out-of-sample forecasting accuracy, combining both traditional univariate seasonal time series methods and novel machine learning and deep learning algorithms. Using empirical data for the United States and Euro Area data the authors conclude that BMA increases the predictive accuracy of CPI inflation forecasts in short-term exercises. Improved forecasting accuracy was also obtained in Bravo et al. (2020, 2021, 2022, 2023) and Ashofteh et al. (2022) and in Bravo and Ashofteh (2023) using Bayesian model ensembles and a metalearning strategy called Arbitrated Dynamic Ensemble (ADE), respectively.

Several studies have already been conducted on forecasting Portugal's GDP using ML. (Garcia, 2020) compared NN with univariate models and found NN to be a more effective forecaster. In addition, Soares (2021) forecasted Portugal's GDP using Multi-Layer Perceptron (MLP) and SVM, focusing on the context of the SARS-CoV-2 pandemic. While there was no direct comparison to econometric methods, the emphasis was still on ML.

Given the advantages of ML, the improved results achieved when using ensemble tree regression models, and the limited number of articles forecasting Portugal's GDP growth using these techniques, this study aims to compare the forecasting effectiveness of ensemble tree regression models for Portugal's GDP with traditional techniques.

In conclusion, the field of macroeconomic forecasting is of utmost importance, providing essential insights for economic policy and decision-making. For instance, Central banks depend on inflation, GDP, employment and money supply forecasts to assess, inform, and regulate standard and non-standard monetary policy instruments (e.g., bank reserve requirement, standing facilities, interest rate policy, asset purchase programs) targeting price stability and long-term economic growth, supporting general economic policy, and setting inflation expectations which enhance policy credibility and efficacy (Bravo & El Mekkaoui, 2022).

Delays in accurate forecasts can lead to less-than-ideal outcomes, highlighting the need for more advanced tools. Recent comparisons between ML models and traditional techniques consistently underscore the superiority of ML, with NN, SVM, and ensemble approaches yielding promising results. The effectiveness of ML, especially when used in an ensemble, presents a strong case for its adoption in the context of GDP growth forecasting, as demonstrated in various studies. This literature review emphasizes the value of ML methods, specifically ensemble tree regression models, in improving the accuracy of GDP growth forecasts and suggests potential for further advancements in the field of macroeconomics.

3. METHODOLOGY

In this section, we will delve into the methodologies used for our forecasting with ML and the methodology utilized by the OECD. The first subsection describes our methodology, detailing the models we used, the metrics for comparison with our benchmark, the specifics of our target, the variables employed, the time periods considered, and the techniques applied. The second subsection focuses on the OECD's methodology, discussing the document where their forecasts are presented, the methods they use, and the variables involved.

3.1. METHODOLOGY OF THE THESIS

To assess whether ML has the potential to surpass traditional forecasting methods, in predicting Portugal's GDP growth, we delve into an examination of four distinct ensemble tree regression models. These models are RF, GBR, extreme gradient boosting (XGB), and light gradient boosting machine (LGBM), which are deployed to generate multiple steps-ahead forecasts. Following this, we juxtapose their forecasts with those produced by the OECD's Economic Outlook. The comparison is facilitated using performance metrics such as RMSE, MPE, and R^2. Our specific focus is on forecasting Portugal's quarterly real GDP growth rate, which is expressed as a percentage change from one period to another, on an annual basis. This rate is adjusted for calendar and seasonal factors and is chained linked to 2016. The forecast periods we concentrate on are the final two quarters of 2019, the final two quarters of 2021, and the entirety of 2022.

When it comes to features, we conducted an extensive analysis of a wide range of macroeconomic quarterly data, in a manner like the OECD's forecast. We considered both "soft" and "hard" indicators (OECD, 2013), and considered the values that were available at the time of the forecast. As a result, our final models were trained using data up until 2007 or 2015[2], a point in time when the values were no longer provisional relative to the forecasts for 2019, 2021, and 2022. Moreover, in terms of the final forecast, we were meticulous in considering the provisional values that were present at that time[3].

---

[2] We created six distinct models for forecasting each successive step. While most of these models were trained on the same number of observations, two of them utilized fewer data points. This discrepancy accounts for the variation in the range of training data, with some models concluding their training data in 2007, while others extended up to 2015.

[3] Forecasting macroeconomic variables presents a challenge due to the provisional nature of the most recent data. Given the complexity of measuring indicators like GDP, these figures are subject to revision over time. Thus, we took care to acknowledge the available data at the time of analysis of OECD's Economic Outlook. Utilizing the current accessible data would introduce bias in favor of ML, as these final values were not yet available during the OECD's forecasts (refer to Appendix A for detailed values and discrepancies regarding Portugal's GDP over time).

Moreover, macroeconomic indicators are not always simultaneously available. For instance, the OECD's Economic Outlook of November 2019 only includes data up to the second quarter of the year in terms of GDP, despite being already in the fourth quarter of 2019. Conversely, certain high-frequency indicators were already

After a comprehensive analysis of the findings and the development of six unique models for each subsequent step, it was determined that the most effective model, which optimizes RMSE, MPE, and R^2, is the GBR across all six steps ahead. The methodology employed to construct these models adhered to the CRISP-DM. This involved the use of tailored techniques for time series forecasting within supervised models. These techniques included methods such as the rolling window and the multiple steps-ahead direct strategy. A more detailed explanation of these techniques will be provided in section 5.

3.2. OECD's Methodology

The OECD generates two annual forecasts, which encompass not only Portugal's GDP growth but also a variety of macroeconomic variables for all member countries. These forecasts are presented in reports known as "Economic Outlook". The creation of these forecasts is based on a combination of model-based analyses, statistical indicator models, and even the judgment of experts and peer review processes (OECD, 2013). In recent years, the OECD has enhanced its forecasting capability by incorporating nowcasting for the current and subsequent quarters. The primary tool for forecasting consists of the National Institute Global Econometric (NiGEM) model. This model is preferred for its reliability over statistical models that utilize high-frequency indicators for near-term forecasting (Turner, 2016). The data that the OECD utilizes comprises both "soft" indicators, such as business sentiment and consumer surveys, and "hard" indicators, which include commodity prices, exchange rates, and industrial production (OECD, 2013).

The forecasting process of the OECD is largely automated, facilitated by a Forecast Entry System that centralizes the necessary data. This system enables experts associated with different countries to verify assumptions, outputs, and new information. It also allows them to adjust necessary parameters to achieve the best possible results. This process takes into consideration both data-driven processes and expert knowledge of each country's current economic and geopolitical environment (OECD, 2013). This combination of data-driven processes and expert knowledge ensures a comprehensive and accurate forecasting process, considering both quantitative data and qualitative insights.

available. We are mindful of these variations and their implications (refer to Appendix B for guidance on accessing the latest historical data relevant to OECD forecasts at the time).

4. DATA

The forecasting exercise is based on a comprehensive set of quarterly macroeconomic data pertaining to Portugal, which spans from the third quarter of 1962 to the second quarter of 2022. The variables are adjusted for calendar effects and seasonality, linked to the 2016 volume, and measured in the national currency[4]. In total, there are nine variables, each with a range of 86 to 227 observations. These nine variables were selected from an initial set of 51 through a rigorous feature selection process (see subsection 5.2) aimed at maximizing model accuracy. Most of these variables were drawn from the OECD's economic outlook database (OECD, 2019), as well as other studies, including Bolhuis & Rayner's (2020) forecast of Turkey's growth and Haque's (2023) forecast of United States GDP growth. The selected variables are:

- Real GDP growth of Portugal.
- Real GDP growth of Spain.
- Real GDP growth of France.
- Real GDP growth of the United Kingdom.
- Real GDP growth of the United States of America.
- Currency spot exchange rate of the US dollar to the national currency of Portugal.
- Final Consumption expenditure of Portugal.
- Recession.
- Year.

These variables stood out primarily due to their significant relationship with the target. In forecasting most time series, particularly those exhibiting strong autocorrelation in the target variable, historical data itself is crucial (Beaumont et al., 1984). Considering that other studies have verified that GDP growth having a strong autocorrelation effect (Arbia & Piras, 2005), we focused on utilizing variables such as the lag of Real GDP growth of Portugal, our target variable, while also considering past recessions and the year to enhance our models' understanding of the temporal evolution of the target.

For the other variables, numerous studies, both empirical and theoretical, including those by Doyle & Faust (2002), demonstrate a robust correlation between GDP growth among countries, which is increasingly strengthened by the rapid globalization of the world economy. This correlation is particularly evident among trading partners, as noted by Jouini (2015), where trade plays a pivotal role in driving this association. Given that Spain, France, the United

[4] There are exceptions to the above-mentioned indications. In terms of chain linked values, OECD utilizes countries' statistics, consequently, Spain's data is linked to 2015 volume, France's to 2014 the United Kingdom's to 2018, and the United States to 2017, rather than being tied to 2016 volume. (OECD ECONOMIC OUTLOOK Database Inventory, 2019).

Kingdom, and the United States (US) are Portugal's major export destinations (OECD, 2017), it's clear why these four variables were prominently featured.

Additionally, Anchelache et al. (2015) identifies final consumption expenditure as a significant predictor of GDP growth in multiple linear regression models. Moreover, exchange rates have been recognized as integral to a country's economic growth, stimulating its economy (Macdonald, 2010).

All the variables mentioned above were sourced from the OECD databank due to their accessibility and to ensure direct comparability with their forecasts. The only exception is the currency spot exchange rate of the US dollar to the national currency of Portugal, which was retrieved from the Federal Reserve Economic Data.

While macroeconomic data is readily available and comes from reliable sources, it poses a significant challenge for ML due to its low frequency. Typically, this data is only available down to a quarterly basis and lacks an extensive historical range. Although most of the variables used in the final forecast date back to the third quarter of 1962, many variables analysed in feature selection, such as final consumption expenditure, the net value of primary income, and many more, start later. This limitation hinders forecasting efforts, especially considering that half of the models being used, namely GBR and RF, cannot handle null values (Bentéjac et al., 2021). Furthermore, ML thrives in data-rich environments (Jean et al., 2016). To address this issue, three configurations of the data were developed: one including variables dating back to the third quarter of 1962, another from the first quarter of 1998, and another from the first quarter of 2001[5]. This will allow us to understand which is the optimal trade-of between observations and the number of variables.

Furthermore, to gain a deeper understanding of how the models respond to the removal of data points, we will conduct a sensitivity analysis on the best models with respect to various historical periods. This analysis focuses on the time periods rather than the existence or absence of variables, as above mentioned, to determine if concept drift should be incorporated into our forecast. Concept drift being the idea that recent data is more relevant for understanding evolving economic dynamics, and removing outdated data points can improve accuracy (Woloszko, 2020).

The historical periods considered start from 1962, the earliest available data point, during Portugal's fascist dictatorship era. We also examine data starting from 1979, a few years after the establishment of the current democratic state. Another key period begins in 1998, when Portugal was well-established in the European Union. Additionally, we consider data from 2002, marking the turn of the new century, the implementation of the euro, and following the

[5] It's important to note that the variables used in the dataset dating back to the third quarter of 1962 are also present in the perspectives of 1998 and 2001. Additionally, the variables from 1998 are included in the 2001 configuration. Data is included in the analysis if it is available.

11th of September of 2001, terrorist attacks in the US. Lastly, we analyse data starting from 2009, in the aftermath of the 2008 economic recession (Eurydice, 2024).

After extensive feature selection and model evaluation, the optimal trade-off between observations and the number of variables led us to choose the above-mentioned features. The insufficiency and low frequency of data concerning Portugal's macroeconomic variables will be further elaborated upon in section 6 of the paper.

Before utilizing the data for modelling, several transformations were performed, including normalization using the standard scaler. While this method isn't ideal due to one of our forecasting objectives being the avoidance of unrealistic assumptions, particularly regarding the stationarity of financial and economic time series, it is necessary because normalization improves machine learning forecasts. The other commonly used methods, such as min-max scaling and decimal scaling, are even less suitable as they require knowledge of the minimum and/or maximum values of a time series, which is impractical for our features and targets since future data may fall outside the range of the training data as indicated by Ogasawara et al. (2019).

Furthermore, we make the conversion of absolute values to their percentage change to facilitate the forecasting of the target, which in this case is the GDP growth rate (for more detailed information regarding the features, please refer to Appendices C and D).

The modelling was developed using Python, with the `Scikit-learn` library being utilized. This library is a popular choice for ML due to its efficiency and versatility, offering a range of supervised learning algorithms.

5. MACHINE LEARNING TECHNIQUES AND MODEL

In this section, we will provide a detailed overview of the procedures, methodologies, and frameworks that were utilized in the development of our models. The section will be divided into several key components. First, we will discuss the handling of outliers, or rather, the decision to not handle them in this context. Next, we will delve into the process of feature selection, where we will elaborate on the methodologies that were employed to determine the variables that were used in the various models. Following this, we will discuss the specific model that was employed in our study. We will then outline the hyper-tuning technique that was applied, which involved the use of a rolling window approach. Finally, we will explain how we forecasted multiple steps. Each of these components plays a crucial role in the overall modelling process, and we will provide a comprehensive explanation of each to provide a clear understanding of our approach.

5.1. OUTLIERS

In the realm of ML, a critical step in the preprocessing phase is the identification and subsequent handling of outliers. These unique data points, despite their potential rarity, can exert a significant influence on forecasting outcomes, often disproportionately compared to other observations. This influence becomes particularly pronounced when dealing with complex datasets, such as those involving macroeconomic variables. Outliers have the potential to skew subsequent observations and can provide valuable insights into the complex interplay between different variables. Furthermore, outliers that may seem highly unlikely could still manifest in the future, thereby affecting the accuracy of forecasts (Bojer, 2022; Rožanec et al., 2021).

In the context of our research, it's important to note that our training data does not include periods that were influenced by unpredictable external economic events such as the COVID-19 pandemic or the conflict in Ukraine. Given the low frequency of our data, we made the conscious decision to retain all observations, choosing not to remove any data points based on them being outliers. This decision was made with the understanding that these outliers could potentially provide valuable insights and contribute to the robustness of our forecasting model.

5.2. FEATURE SELECTION

In the context of feature selection, we meticulously crafted a comprehensive process. This process was multifaceted, involving the evaluation of feature importance outcomes derived from a variety of methods, the integration of these outcomes, the complete disregard of these outcomes, and the exclusive consideration of the lagged target. This intricate approach was implemented across three separate datasets, each characterized by different quantities of observations and features, as described in the sixth paragraph of section 4. The need for such differentiation arose from the diverse relationships that exist between features, especially when fewer variables or observations come into play (Jović et al., 2015).

The methods we employed in our approach included Spearman's correlation, embedded techniques, and wrapper methods that leverage decision trees. These methods were chosen over alternatives such as recursive feature selection, Lasso embedded methods, and dimensionality reduction techniques. The primary reason for this selection was that the latter group of methods relies on linear relationships between variable (Jović et al., 2015).

The main driving force behind our decision to use ML over econometrics was ML's ability to circumvent the assumptions that are deeply ingrained in traditional statistical methods. Therefore, if we were to retain techniques that are centred around linear relationships, it would be in direct contradiction with the fundamental objective of this thesis. This objective is to explore and leverage the power of ML in understanding and interpreting complex relationships between variables, which traditional statistical methods might not be able to fully capture due to their inherent assumptions (Kaushik, 2020).

In essence, our feature selection process was designed to be robust and flexible, capable of adapting to the unique characteristics of each dataset and the complex relationships between their features. This approach allows us to fully harness the potential of ML in our research.

5.3. GRADIENT BOOSTING REGRESSOR

The GBR is a supervised ML model that is specifically designed for regression tasks. It operates using a sequential ensemble approach, which involves the utilization of base-learners to iteratively refine its predictions.

The unique strength of GBR lies in its ability to enhance its forecasts with each iteration by learning from the errors of previous predictions. This iterative learning process allows GBR to continuously improve its predictive accuracy, making it a powerful tool for regression tasks. This process of learning from past errors and refining future predictions is a defining characteristic of GBR and is what distinguishes it from other ensemble models (Natekin & Knoll, 2013).

Models like RF, while also being ensemble models, do not possess this iterative learning capability. Instead, they rely on the simple aggregation of the outputs of their base-learners, without any mechanism for learning from past errors or refining future predictions. This difference in approach is what sets GBR apart and contributes to its superior performance in regression tasks (Natekin & Knoll, 2013).

The equation below illustrates the sum of estimation functions, the key component described above. This equation represents the cumulative effect of the estimations, which are refined iteratively to improve the model's predictions. However, it's important to note that this equation does not yet show the sum of the base-learners, which is another crucial part of the GBR model's operation.

$$\hat{f}(x) = \sum_{i=0}^{M} \widehat{f_i}(x), \tag{1}$$

Where $\hat{f}(x)$ denotes the estimation function, M is number of iterations, and the first iteration is the initial guess of the model.

Building on the previous discussion, the GBR model's operation is underpinned by several key components. These include the loss function, the base learner, and the number of iterations. The loss function serves as a metric, quantifying the disparity between the predictions generated by the model and the actual data. The base learners, which are incorporated iteratively manner to enhance the model's forecasts, can be constructed using a variety of methods, including decision trees and splines, among others (Natekin & Knoll, 2013).

This iterative learning process is encapsulated in the equation that follows. This equation demonstrates how each iteration of the model is computed, with each new prediction being the sum of the previous estimation function and the product of the base learner and the learning rate. This equation, therefore, provides a mathematical representation of the GBR model's iterative learning process, highlighting how it continuously refines its predictions to improve accuracy (Natekin & Knoll, 2013).

$$\widehat{f_i} = \widehat{f_{i-1}} + \rho_i h(x, \theta_i), \tag{2}$$

Where $\widehat{f_i}$ denotes the estimation function for the iteration i, $\widehat{f_{i-1}}$ the result of the previous iteration of the estimation function, ρ_i being the learning rate associated to the base-learner of said iteration, and $h(x, \theta_i)$ the base learner of the iteration.

The operational process of the GBR algorithm can be broken down into the following steps:

1. Initialization: The estimation function is initialized with a constant value, typically the mean of the target variable.
2. Iteration Loop:
   a. Calculate Negative Gradient: The negative gradient of the loss function with respect to the predicted values is computed at each iteration.
   b. Fit a New Base Learner: A new base learner is fitted to the negative gradient calculated in the previous step.
   c. Update Estimation Function: The estimation function is updated by adding the prediction of the new base learner, which is multiplied by a learning rate, to the previous estimation function.

GBR stands out as an ensemble model that is highly proficient at capturing complex non-linear functions. Its effectiveness in forecasting real-world data has been extensively documented. One of the key strengths of GBR is its exceptional flexibility in customization, particularly in the selection of loss functions and base learners. However, this versatility comes with a trade-off in terms of computational cost. Due to its iterative optimization approach, GBR is more computationally intensive compared to other ensemble models such as RF (Natekin & Knoll, 2013).

In our forecasting model, we utilized the mean squared error loss function. Although this is not one of our primary metrics, it assists in optimizing the RMSE and R^2, while also penalizing significant errors. Furthermore, we chose decision trees as our base learners. This choice was made because decision trees do not rely on assumptions of linearity, they contribute to the diversity of the model, and they provide computational efficiency (Chicco et al., 2021).

5.4. ROLLING WINDOW TECHNIQUE

When it comes to fine-tuning parameters for ML models, there are various approaches. In this subsection, we will start by describing one of the simplest methods: the train, validation, and test method. This will be followed by an explanation of the cross-validation method. Finally, we will discuss the rolling window technique, including its two variants, which is the method we utilized.

In the process of enhancing the performance of a model via hyperparameter tuning, a prevalent methodology involves the division of the dataset into three distinct segments: training, validation, and test sets (refer to Figure 1). The training data serves the purpose of fitting the model, utilizing the parameters that have been defined. On the other hand, the validation set is exclusively set aside for the task of choosing the most effective model, thereby avoiding any bias that could potentially be introduced if the test set was used for this purpose. The final segment, the test set, is then employed to assess the performance of the selected model when it is exposed to data that it has not encountered before (Fi & Garriga, 2010). This approach ensures a robust and unbiased evaluation of the model's ability to generalize from the training data to unseen data, which is a critical aspect of model performance in practical applications.

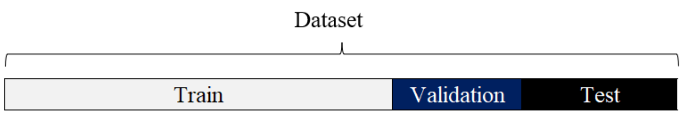

Figure 1 – Train, validation, test split.

In circumstances where the available data is scarce, such as in our case, the traditional approach of dividing the data into training, validation, and test sets can result in a substantial reduction of training data. To counteract this issue, the technique of k-fold cross-validation is frequently utilized. This method deviates from the rigid partitioning into training, validation, and test sets. Instead, it involves the division of the training data into k-folds, which are generally subsets of equal size. Each of these folds is subsequently used as a validation set, while the model is trained using the remaining k-1 folds. This procedure is carried out for each fold, ensuring that every data point is used for both training and validation purposes (refer to Figure 2). Consequently, this allows for a more comprehensive assessment of the model's performance, without the need to forfeit valuable data (Fi & Garriga, 2010).

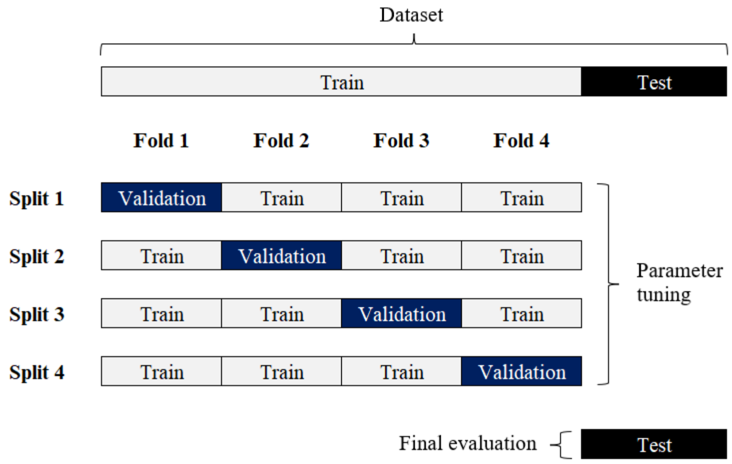

Figure 2 – Cross-validation technique.

The challenge with using cross-validation for time series forecasting is that it overlooks the temporal nature of the data. As a result, a different strategy, referred to as the rolling window method, is required. In this method, the k-folds are arranged before the validation set to ensure that the validation is not influenced by any future data. As each fold is processed, both the training and validation sets advance in a chronological order.

The rolling window method comes in two forms: anchored and non-anchored. In the anchored variant, the validation set moves forward in tandem with the training set, incorporating both the previous training data and the past validation data from the preceding folds (refer to Figure 3). On the other hand, the non-anchored variant does not accumulate training data but progresses similarly to the validation set (refer to Figure 4).

The non-anchored method has certain advantages over the anchored one, particularly in terms of reducing bias during the tuning of hyperparameters, especially when dealing with past data that has been influenced by the changing complexity of macroeconomic variables. When compared to other cross-validation methods, this approach stands out due to its robustness, its focus on temporal dynamics, and its realistic training for future data (Bojer, 2022). In our modelling efforts, we employ both variants of the rolling window method.

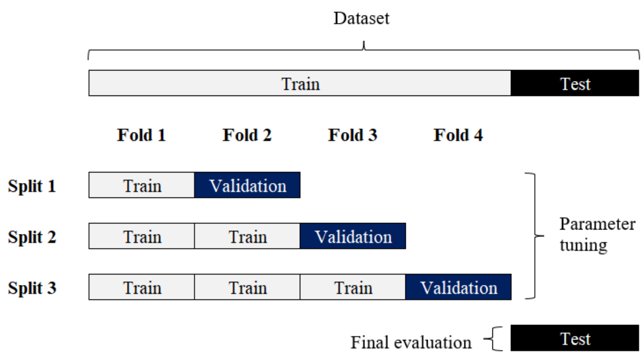

Figure 3 – Anchored rolling window approach.

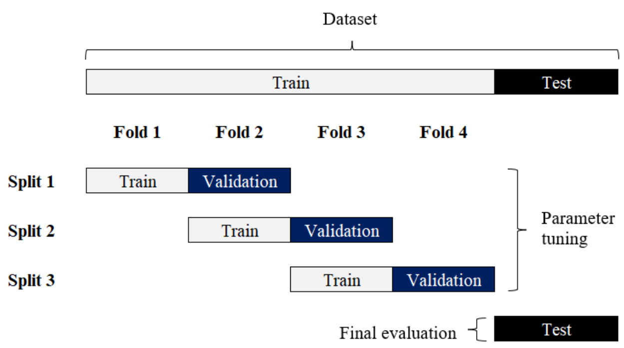

Figure 4 – Non-anchored rolling window approach.

5.5. Multiple steps ahead approach

When forecasting GDP growth for up to 6 quarters ahead using the latest available data, it is crucial to adopt a multi-step strategy, especially when information is limited. Several strategies are designed for this purpose: the recursive strategy, the direct strategy, the multi-input multi-output (MIMO) strategy, and the DIRMO strategy. In this subsection of the thesis, we will briefly analyse how each of these four methods works, highlighting their advantages and disadvantages. Additionally, we will indicate the chosen strategy for our models and explain our choice.

The recursive strategy, also referred to as the iterated method, is the oldest technique among the four discussed. It employs the predicted data from one step ahead as input for the subsequent prediction (Bontempi et al., 2013). This approach falls within the category of single-output methods, as it uses multiple inputs to produce a single output. Mathematically, this can be represented by the following equation:

$$\varphi_{t+1} = f(\varphi_t, \varphi_{t-1}, \dots, \varphi_{t-d+1}) + w, \tag{3}$$

Where φ_{t+1} denotes the value of the next step, f is the function that calculates φ_{t+1} based on its lag values ranging from $t-d$ to t, d being the number of past values considered in the forecast, and w being a scalar noise term that averages to 0. Once the estimation function is established, the forecast for one to three steps ahead is handled as follows:

$$\hat{\varphi}_{N+h} = \begin{cases} \hat{f}(\varphi_N, \dots, \varphi_{N-d+1}) & \text{if } h = 1 \\ \hat{f}(\varphi_{N+h-1}, \dots, \varphi_{N+1}, \varphi_N, \dots, \varphi_{N-d+h}) & \text{if } h \in \{2, \dots, d\}, \\ \hat{f}(\varphi_{N+h-1}, \dots, \varphi_{N+h-d}) & \text{if } h \in \{d+1, \dots, H\} \end{cases} \tag{4}$$

Although this strategy is simple, easy to implement, and provides additional data for subsequent steps, its main disadvantage is the accumulation of prediction errors. This leads to increasingly inaccurate forecasts the further they extend beyond the last available value, especially if the previous forecasts are not very accurate (Taieb et al., 2010).

Moving on to the next strategy within the family of single-output methods, the direct strategy. In this approach, we develop individual models to forecast each step, training each model with different sets of lagged data to compensate for the information deficit (Bontempi et al., 2013). This can be summarized by the following equations:

$$\varphi_{t+h} = f_h(\varphi_t, \varphi_{t-1}, \dots, \varphi_{t-d+1}) + w, \qquad \text{with } h \in \{1, \dots, H\}, \tag{5}$$

For the most part, this equation works in the same way as the first equation of the recursive strategy. However, instead of calculating just the step ahead of t , it can calculate any step ahead from 1 to H depending on the step ahead we are targeting and, consequently, which model we are utilizing. The estimation function for this strategy looks like this:

$$\widehat{\varphi}_{N+h} = \hat{f}_h(\varphi_N, \dots, \varphi_{N-d+1}), \tag{6}$$

The main advantage of the direct strategy is that it avoids the accumulation of errors by not using the forecast of the model as input for subsequent steps. However, this method has the drawback of developing individual models independently for each step, which means it cannot account for the complex dependencies between the variables at different steps. Additionally, it is computationally expensive because it requires creating a separate model for each forecast step (Taieb et al., 2010).

To bypass these issues, we can consider the other family of multiple-step-ahead techniques, known as multiple-outputs. The MIMO strategy uses a single model to forecast multiple steps simultaneously (Bontempi et al., 2013). The equation that represents the desired result in this strategy is:

$$\{\varphi_{t+H}, \dots, \varphi_{t+1}\} = F(\varphi_t, \varphi_{t-1}, \dots, \varphi_{t-d+1}) + w, \tag{7}$$

Where φ_{t+1} to φ_{t+H} are the predicted values at time steps $t+1$ to $t+H$, where H represents the forecast horizon. F represents the forecasting function and w the residual.

The equation functions similarly to the others, but in this technique, it forecasts multiple steps in a single go, and w is no longer a scalar but a vector. The estimation function looks like this:

$$\{\widehat{\varphi}_{N+H}, \dots, \widehat{\varphi}_{N+1}\} = \hat{F}(\varphi_N, \varphi_{N-1}, \dots, \varphi_{N-d+1}), \tag{8}$$

The main difference between this estimation function and the others is that the forecast is no longer a scalar, or a single value, but a time series. While this method avoids the disadvantages of the previous two methods, it constrains the model to fixed parameters that forecast the entire series rather than focusing on individual moments. This makes it less flexible than the other two strategies.

To conclude this subsection, we look at the second strategy within the multiple-outputs family, DIRMO. This hybrid approach divides the forecast steps into sections, initially resembling the MIMO strategy and then transitioning to a recursive model. It uses the results of the first forecast section as input data while also creating separate models for each section, like the direct strategy (Bontempi et al., 2013). The equation for this approach looks like this:

$$\left\{\varphi_{t+ps}, \dots, \varphi_{t+(p-1)s+1}\right\} = F_p(\varphi_t, \varphi_{t-1}, \dots, \varphi_{t-d+1}) + w, \qquad p \in \{1, \dots, n\}, \tag{9}$$

Where $\varphi_{t+(p-1)s+1}$ to φ_{t+ps} are the predicted values at time steps $t + (p-1)s + 1$, where s represents the spacing or stride between predictions, and p indicates the prediction step or horizon. The set $\{1, \dots, n\}$ defines the range of steps being considered. F_p represents the forecasting function for the pth step ahead. w represents the residual for the pth step ahead.

The equation is quite like that of the MIMO strategy, with the main difference being that s represents the number of sections forecasted, each with multiple outputs. The equation function looks like this:

$$\left\{\widehat{\varphi}_{N+ps}, \dots, \widehat{\varphi}_{N+(p-1)s+1}\right\} = \hat{F}_p(\varphi_N, \varphi_{N-1}, \dots, \varphi_{N-d+1}) + w, \qquad p \in \{1, \dots, n\}, \tag{10}$$

This strategy is not limited to a single model and fixed parameters, thus avoiding the inflexibility of MIMO. It can interpret the complexity of relationships within a time series, unlike the direct method, and it mitigates error accumulation since the first section of multiple outputs is not based on previous forecasted data. The main drawback of this strategy is the challenge of selecting the optimal number of sections s for dividing the forecast a priori (Taieb et al., 2010).

In our methodology, we have opted for the direct strategy because it is straightforward to understand and implement, avoids the accumulation of prediction errors, and crucially, allows us to determine the most effective models for forecasting our target at each step. This approach offers a practical solution for accurate forecasting. It involves training six separate models using the available lagged data at each step (Bontempi et al., 2013).

6. RESULTS

After implementing all above-described methods and techniques we have verified that for the 10 instances being analysed, these being from the last two quarters of 2019, the last two quarters of 2021, and the entirety of 2022 (with the last two quarters forecasted twice) OECD's results showed an RMSE of 5.84, an MPE of -95%, and an R^2 of 0.09, while ML yielded an RMSE of 10.02, an MPE of -163%, and an R^2 of -0.56.

While OECD's overall forecasts outperformed ML, there were instances where ML demonstrated better performance across all three metrics. Specifically, in the last two quarters of 2019 and 2022 when employing one-step and two-steps ahead forecasting. OECD reported an RMSE of 1.54, an MPE of 46%, and an R^2 of -4.05, whereas ML achieved an RMSE of 0.33, an MPE of 20%, and an R^2 of -0.07. Additionally, ML outperformed OECD in terms of RMSE and R^2 for the 2022 forecast when utilizing three to six steps-ahead. OECD's RMSE values were 3.61 of with an R^2 of -0.004, while ML achieved an RMSE of 2.31 and an R^2 of 0.59. However, it's worth noting that for 2022, OECD did present a superior MPE of -283% compared to ML's MPE of -449% (see Figure 5 and Table 1).

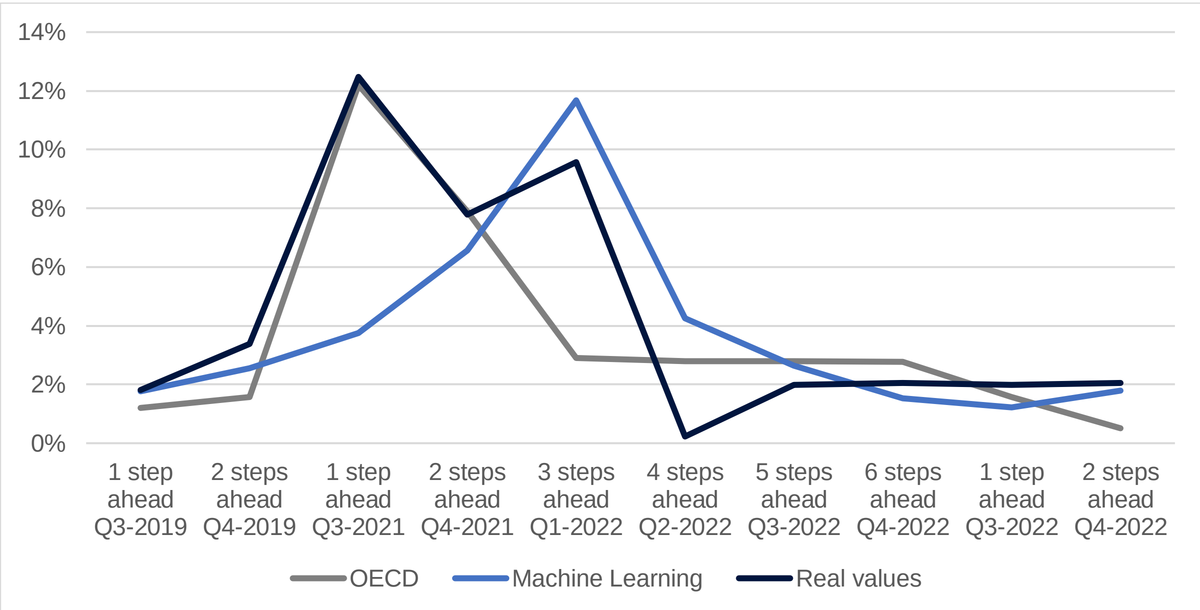

Figure 5 – Linear graph of OECD's and Machine Learnings forecasts of real GDP growth of Portugal compared with real values.

Note: the real values of Portugal's real GDP growth were retrieved from the Economic Outlook of December 2023 of OECD.

Table 1 – Table of metrics of RMSE, MPE, and R^2 of OECD's and Machine Learnings forecasts of real GDP growth of Portugal compared with real values.

Forecast	Metric	All instances (10 instances)	1-2 steps-ahead Q3-Q4 2019 & Q3-Q4 2022 (4 instances)	3-6 steps-ahead Q1-Q4 2022 (4 instances)
OECD	RMSE	5.8373	1.5432	3.6095
	MPE	-94.58%	45.99%	-282.65%
	R^2	0.0891	-4.0518	-0.0041
ML	RMSE	10.0153	0.3279	2.3132
	MPE	-163.14%	19.59%	-448.86%
	R^2	-0.5628	-0.0735	0.5876

Note: the real values of Portugal's real GDP growth were retrieved from the Economic Outlook of December 2023 of OECD.

Upon examination, it was found that each of the six individual models, trained for step-by-step forecasting from one to six steps ahead, achieved their highest accuracy using the GBR model. Additionally, four out of the six models used data starting from 1962. while two models used data starting from 1998 when only considering the time periods of 1962, 1998, and 2001. It was also verified that four of the six models employed a non-anchored method for the rolling window technique, whereas the remaining two used the anchored method. Regarding the model parameters, no specific pattern was observed, as each individual model predominantly used unique sets to maximize their accuracy in testing (see Appendix E).

Regarding the sensitivity analysis across the time periods of 1962, 1979, 1998, 2002, and 2009 for the best models and respective parameters in each step-ahead, we observe that for the first step-ahead, the best metrics were 0.60 for RMSE, -7% for MPE, and 0.29 for R^2, starting from 1962. For the second step-ahead, the best metrics were 0.31 for RMSE, -55% for MPE, and -0.47 for R^2, starting from 1998. For the third step-ahead, the best metrics were 0.76 for RMSE, 14% for MPE, and -0.12 for R^2, starting from 1962. For the fourth step-ahead, the best metrics were 0.85 for RMSE, -8% for MPE, and -0.42 for R^2, starting from 1962. For the fifth step-ahead, the best metrics were 0.80 for RMSE, 12% for MPE, and -0.25 for R^2, also starting from 1962. For the sixth step-ahead, the best metrics were 0.81 for RMSE, -30% for MPE, and -0.28 for R^2, starting from 1998 (see Appendix F). We can verify that the optimal time periods are the same as when analysing with 1962, 1998, and 2001, as seen above.

Although it was verified that for most steps the best time period to consider is starting from 1962, we must consider that no single time span maximizes all three metrics in test for any of the steps-ahead. For instance, in the first step, starting from the most recent period, that being 2009, we achieve the lowest RMSE with a value of 0.25. The best MPE is with the data starting in 1962 with -7%, and for R^2, it is starting from 1979, with a value of 0.36. We arrived at the above-mentioned conclusions using the same method developed to find the optimal parameters and best model for the original forecast (see Appendix G). If we were to only consider RMSE, 2009 would be the best starting point for all the steps, showcasing how much the evaluation method can affect the conclusions of analysis and showing how paramount it's to have an array of metrics to analyse (see Appendix F).

Given the outcomes discussed earlier we can now discuss an array of conclusions and questions. We will start off with the most prominent question regarding the final forecast and that will affect directly if our hypothesis was confirmed, how did the OECD manage to outperform ML so significantly in the overall assessment? The discrepancy stems from the predictions for the third quarter of 2021 when employing one step-ahead forecast. While the OECD nearly achieved a perfect forecast, displaying a RMSE of 0.27 and a MPE of 2%, the ML model's forecasts fell short. The ML model exhibited significant underperformance with a RMSE of 8.71 and a MPE of 70%.

So, after these results, we must consider why exactly ML was unable to forecast the third quarter of 2021 in such an effective way as the ones of 2019 and 2022 relative to OECD's forecast.

In addressing the initial inquiry, it's imperative to acknowledge that ML excels in data-rich settings. Our forecasting relies on macroeconomic data available only quarterly, which can hinder its forecasting accuracy compared to more conventional methods like those employed by the OECD (Lim & Zohren, 2021). In certain cases, ML forecasts may even perform worse than simpler models (Makridakis et al., 2018). Moreover, ML outcomes are highly sensitive to available data; various studies indicate that the same model can outperform econometric methods in specific scenarios while underperforming in others, depending on the dataset utilized (Bojer, 2022; Woloszko, 2020). It's worth mentioning that while ML was not able to outperform those of the OECD, as evidenced in our results (see graph 1 and table 1), they were significantly penalized for their inaccurate prediction of the first quarter of 2021 relative to the OECD's forecast.

Regarding the other conclusions, in our analysis of various forecasting models, it was consistently evident that GBR outperformed other models across all forecast horizons in terms of RMSE, MPE, and R^2 metrics. Additionally, it was verified after sensitive analysis that, for the most part, the models performed best when utilizing the greatest number of instances as possible, using data starting from 1962 and the use of non-anchored rolling window technique. These observations underscore the significance of concept drift by prioritising a higher number of instances over only considering more recent data. Such happens due to ML models exhibiting a preference for higher frequency data, as lowering the number of observations further deteriorates forecast accuracy given the already low frequency of economic data.

Despite the advantage of ML in identifying complex patterns without relying on linear assumptions, our extensive feature selection process reduced the initial pool of 51 variables to just a few essential ones. Most of the variables used demonstrated a strong linear relationship with the target, as evidenced by Spearman’s correlation (check Appendix H). Other studies, such as Woloszko (2020), have reached similar conclusions, noting that identical ML models perform better with variables that exhibit linear relationships with the target.

In essence, the thesis did not confirm its hypothesis of "Can ML techniques, specifically ensembles of regression trees, arrive at better results than OECD’s forecast for the GDP growth of Portugal?” by verifying that ML did not outperform OECD's forecasts in all three metrics (RMSE, MPE, and R^2) for ten different forecasts spanning the last two quarters of 2019, the last two quarters of 2021, and the entirety of 2022 (with the last two quarters forecasted twice). While ML often surpassed OECD predictions in various cases, it generally underestimated overall economic growth, consistently displaying lower MPE values compared to OECD forecasts. Additionally, ML forecasts occasionally diverged significantly from actual values, notably observed in the third quarter of 2021. Hence, it's crucial to subject ML forecasting techniques to thorough evaluation against established benchmarks such as OECD before widespread adoption in economic institutions, as advocated in recent literature (Bojer, 2022).

7. CONCLUSIONS

The objective of this thesis was to evaluate whether ML could surpass forecasts produced by established institutions in forecasting Portugal’s GDP growth using ensemble regression tree models whilst using OECD’s economic outlook as a benchmark. The results reveal a complex picture of ML's strengths and limitations in economic forecasting.

The primary findings indicate that while ML achieved superior forecasts compared to the OECD in most cases, it ultimately underperformed due to one significant outlier that deviated substantially from the actual value. This outlier underscores the sensitivity of ML models to specific data points and their potential for larger errors in certain scenarios, suggesting that ML is, overall, less reliable than forecasts produced by established institutions.

Several factors contributed to these results. ML models thrive in data-rich environments and perform better with high frequency data. However, the reliance on quarterly macroeconomic data constrained their performance. Additionally, although ML can capture complex patterns, the best-performing models were those with fewer variables, typically exhibiting strong linear relationships with the target, thus diminishing ML's advantages.

Furthermore, the study found that GBR was the most effective model for forecasting Portugal’s GDP growth across all time horizons. It also showcased a preference for more observation points over a larger number of variables suggesting that ML models value historical data even when it appears less relevant for future forecasts.

In conclusion, while ML models have shown potential in specific instances, their overall performance in this study was not better than forecasts produced by established institutions for GDP growth forecasting. The findings suggest that ML should be viewed as a complementary tool rather than a substitute for established techniques. The mixed results highlight the importance of benchmarking ML forecasts against traditional methods before considering their broader application in economic forecasting. Future research should explore integrating ML with forecasts produced by established institutions to leverage the strengths of both approaches, potentially leading to more robust and accurate forecasts.

BIBLIOGRAPHICAL REFERENCES

Ashofteh, A., Bravo, J. M., & Ayuso, M. (2022). A New Ensemble Learning Strategy for Panel Time-Series Forecasting with Applications to Tracking Respiratory Disease Excess Mortality during the COVID-19 pandemic. Applied Soft Computing. Volume 128, Article 109422.

Anchelache, C., Manole, A., & Anghel, M. (2015). *Analysis of final consumption and gross investment influence on GDP – multiple linear regression model*. http://www.icaap.org

Arbia, G., & Piras, G. (2005). *Convergence in per-capita GDP across European regions using panel data models extended to spatial autocorrelation effects*. www.isae.it

Beaumont, C., Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1984). Forecasting: Methods and Applications. *The Journal of the Operational Research Society*, *35*(1), 79. https://doi.org/10.2307/2581936

Ben Taieb, S., Sorjamaa, A., & Bontempi, G. (2010). Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing*, *73*(10–12), 1950–1957. https://doi.org/10.1016/j.neucom.2009.11.030

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*(3), 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Bojer, C. S. (2022a). Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities. *International Journal of Forecasting*, *38*(4), 1555–1561. https://doi.org/10.1016/j.ijforecast.2021.11.003

Bojer, C. S. (2022b). Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities. *International Journal of Forecasting*, *38*(4), 1555–1561. https://doi.org/10.1016/j.ijforecast.2021.11.003

Bolhuis, M. A., & Rayner, B. (2020). *Deus ex Machina? A Framework for Macro Forecasting with Machine Learning, WP/20/45, February 2020*.

Bontempi, G., Ben Taieb, S., & Le Borgne, Y. A. (2013). Machine learning strategies for time series forecasting. *Lecture Notes in Business Information Processing*, *138 LNBIP*, 62–77. https://doi.org/10.1007/978-3-642-36318-4_3

Bravo, J. M. (2021). Forecasting Longevity for Financial Applications: A First Experiment with Deep Learning Methods. In: Kamp M. et al. (eds) Machine Learning and Principles and Practice of Knowledge Discovery in Databases. ECML PKDD 2021. Communications in Computer and Information Science, vol 1525. Springer, Cham. pp. 232–249. https://doi.org/10.1007/978-3-030-93733-1_17.

Bravo, J. M. (2021). Forecasting mortality rates with Recurrent Neural Networks: A preliminary investigation using Portuguese data. CAPSI 2021 Proceedings 7 (Atas da 21ª Conferência da Associação Portuguesa de Sistemas de Informação 2021). https://aisel.aisnet.org/capsi2021/7/

Bravo, J. M. (2022). Pricing Participating Longevity-Linked Life Annuities: A Bayesian Model Ensemble approach. European Actuarial Journal, 12, 125–159.

Bravo, J. M., & Ashofteh, A. (2023). Ensemble Methods for Consumer Price Inflation Forecasting. CAPSI 2023 - 23rd Conference of the Portuguese Association for Information Systems (23ª Conferência da Associação Portuguesa de Sistemas de Informação), pp. 317–336. DOI: 10.18803/capsi.v23.317-336.

Bravo, J. M., & Ayuso, M. (2021) Forecasting the Retirement Age: A Bayesian Model Ensemble Approach. In: Rocha Á., Adeli H., Dzemyda G., Moreira F., Ramalho Correia A.M. (eds) Trends and Applications in Information Systems and Technologies. WorldCIST 2021. Advances in Intelligent Systems and Computing, Volume 1365 AIST, pp. 123 – 135. Springer, Cham.

Bravo, J. M., & Ayuso, M. (2021). Linking Pensions to Life Expectancy: Tackling Conceptual Uncertainty through Bayesian Model Averaging. Mathematics, 9(24): 3307, p. 1–27.

Bravo, J. M., & Santos, V. (2021). Backtesting Recurrent Neural Networks with Gated Recurrent Unit: Probing with Chilean Mortality Data. In: Garcia, M.V., Fernández-Peña, F., Gordón-Gallegos, C. (eds) Advances and Applications in Computer Science, Electronics, and Industrial Engineering. CSEI 2021. Lecture Notes in Networks and Systems, vol 433, Springer, Cham. pp. 159–174. https://doi.org/10.1007/978-3-030-97719-1_9.

Bravo, J. M., Ayuso, M. (2020). [Mortality and life expectancy forecasts using bayesian model combinations: An application to the portuguese population] Previsões de mortalidade e de esperança de vida mediante combinação Bayesiana de modelos: Uma aplicação à população portuguesa. RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao, E40, 128–144.

Bravo, J. M., Ayuso, M., Holzmann, R. & Palmer, E. (2021). Addressing the Life Expectancy Gap in Pension Policy. Insurance: Mathematics and Economics, 99, 200-221.

Bravo, J. M., Ayuso, M., Holzmann, R., and Palmer, E. (2023). Intergenerational Actuarial Fairness when Longevity Increases: Amending the Retirement Age. Insurance: Mathematics and Economics 113, 161-184.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, 1–24. https://doi.org/10.7717/PEERJ-CS.623

Doyle, B. M., & Faust, J. (2002). *An Investigation of Co-movements among the Growth Rates of the G-7 Countries*.

Eurydice. (2024). *Historical development Portugal's history*.

Fi, M. O., & Garriga, G. C. (2010). Permutation Tests for Studying Classifier Performance Markus Ojala. In *Journal of Machine Learning Research* (Vol. 11).

Garcia J. (2020). *Forecasting Portuguese GDP A comparison of univariate time series models*.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2020). *How is Machine Learning Useful for Macroeconomic Forecasting? **.

Haque, S. (2023). *Journal of Mathematics and Statistics Studies Retail Demand Forecasting Using Neural Networks and Macroeconomic Variables*. https://doi.org/10.32996/jmss

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). *Combining satellite imagery and machine learning to predict poverty*. http://science.sciencemag.org/

Jouini, J. (2015). Linkage between international trade and economic growth in GCC countries: Empirical evidence from PMG estimation approach. *Journal of International Trade and Economic Development*, *24*(3), 341–372. https://doi.org/10.1080/09638199.2014.904394

Jović, A., Brkić, K., & Bogunović, N. (2015). *A review of feature selection methods with applications*.

Jung, J.-K., Patnam, M., & Ter-Martirosyan, A. (2018). *An Algorithmic Crystal Ball: Forecasts-based on Machine Learning, WP/18/230, November 2018*.

Kaushik, M. (2020). *Forecasting Foreign Exchange Rate: A Multivariate Comparative Analysis between Traditional Econometric, Contemporary Machine Learning & Deep Learning Techniques I. I NTRODUCTION*.

Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (Vol. 379, Issue 2194). Royal Society Publishing. https://doi.org/10.1098/rsta.2020.0209

Macdonald, R. (2010). *THE ROLE OF THE EXCHANGE RATE IN ECONOMIC GROWTH: A EURO-ZONE PERSPECTIVE NATIONAL BANK OF BELGIUM WORKING PAPERS-RESEARCH SERIES*. http://www.nbb.be

Maehashi, K., & Shintani, M. (2020). Macroeconomic Forecasting Using Factor Models and Machine Learning: An Application to Japan. *Journal of the Japanese and International Economies*, *58*. https://doi.org/10.1016/j.jjie.2020.101104

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, *13*(3). https://doi.org/10.1371/journal.pone.0194889

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*(DEC). https://doi.org/10.3389/fnbot.2013.00021

OECD. (2013). *Forecasting methods and analytical tools*. https://doi.org/10.1787/275257320252

OECD. (2017). *PORTUGAL TRADE AND INVESTMENT STATISTICAL NOTE*. www.oecd.org/investment/trade-

OECD. (2019). *OECD ECONOMIC OUTLOOK Database Inventory*. http://www.oecd.org/eco/outlook/economicoutlook.htm

Ogasawara, E., Martinez Leonardo, Oliveira Daniel, Zimbrão Geraldo, & Mattoso Gisele. (2019). *Adaptive Normalization: A Novel Data Normalization Approach for Non-Stationary Time Series*.

Rožanec, J., Trajkova, E., Kenda, K., Fortuna, B., & Mladenić, D. (2021). Explaining bad forecasts in global time series models. *Applied Sciences (Switzerland)*, *11*(19). https://doi.org/10.3390/app11199243

Soares, I. C. (2021). *Prediction of the quarterly GDP in Portugal: The application of the machine learning tools in presence of SARS-CoV-2 pandemic*.

Soybilgen, B., & Yazgan, E. (2021). Nowcasting US GDP Using Tree-Based Ensemble Models and Dynamic Factors. *Computational Economics*, *57*(1), 387–417. https://doi.org/10.1007/s10614-020-10083-5

The World Bank. (2022). *World Development Report 2022*. https://www.worldbank.org/en/publication/wdr2022/brief/chapter-1-introduction-the-economic-impacts-of-the-covid-19-crisis

Turner, D. (2016). *THE USE OF MODELS IN PRODUCING OECD MACROECONOMIC FORECASTS*. www.oecd.org/eco/workingpapers

U.S. Cemters for Disease Control and Prevention. (2019). *CDC Museum COVID-19 Timeline*. https://www.cdc.gov/museum/timeline/covid19.html

Woloszko N. (2020). *ADAPTIVE TREES-A NEW APPROACH TO ECONOMIC FORECASTING*.

Yoon, J. (2021). Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. *Computational Economics*, *57*(1), 247–265. https://doi.org/10.1007/s10614-020-10054-w

APPENDIX A

Table with the data available at the time of the issue regarding real GDP growth rate of Portugal of OECD from period to period, expressed annually.

Date	November 2019	December 2021	December 2022	November 2023	Difference between November 2023 and values used for the forecast
Q1-2019	2.33%	3.57%	3.58%	3.52%	1.19%
Q2-2019	2.30%	2.28%	2.30%	2.37%	0.06%
Q3-2019	-	1.86%	1.82%	1.82%	-
Q4-2019	-	3.34%	3.32%	3.38%	-
Q1-2020	-	-16.52%	-16.32%	-16.40%	-
Q2-2020	-	-48.41%	-48.09%	-48.06%	-
Q3-2020	-	72.92%	72.41%	72.50%	-
Q4-2020	-	1.10%	1.49%	1.41%	-
Q1-2021	-	-12.48%	-9.89%	-9.39%	3.09%
Q2-2021	-	19.32%	18.91%	19.03%	-0.29%
Q3-2021	-	-	11.60%	12.47%	-
Q4-2021	-	-	7.98%	7.79%	-
Q1-2022	-	-	9.88%	9.57%	-0.31%
Q2-2022	-	-	0.51%	0.23%	-0.28%
Q3-2022	-	-	-	1.98%	-
Q4-2022	-	-	-	2.05%	-

Note: The only values used for the forecast were the two lags beforehand, therefore, the only relevant values for the forecast are the first and second quarters of 2019, 2021, and 2022.

APPENDIX B

Table with the last available data for OECD's forecast of each Economic Outlook by group of variables.

OECD's Economic Outlook	Years forecasted	Trimesters forecast	Variable group	Last historical point
December 2021	2021-2022	2021-Q3 to 2022-Q4	Prices	2021-Q3
December 2021	2021-2022	2021-Q3 to 2022-Q4	Employment	2021-Q3
December 2021	2021-2022	2021-Q3 to 2022-Q4	Exchange rate	2021-Q3
December 2021	2021-2022	2021-Q3 to 2022-Q4	Interest rates	2021-Q3
December 2021	2021-2022	2021-Q3 to 2022-Q4	Other	2021-Q2
November 2019	2019	2019-Q3 to 2019-Q4	Prices	2021-Q3
November 2019	2019	2019-Q3 to 2019-Q4	Employment	2021-Q3
November 2019	2019	2019-Q3 to 2019-Q4	Exchange rate	2021-Q3
November 2019	2019	2019-Q3 to 2019-Q4	Interest rates	2021-Q3
November 2019	2019	2019-Q3 to 2019-Q4	Other	2021-Q2
December 2020	2020-2021	2021-Q1 to 2021-Q4	Prices	2020-Q3
December 2020	2020-2021	2021-Q1 to 2021-Q4	Employment	2020-Q3
December 2020	2020-2021	2021-Q1 to 2021-Q4	Exchange rate	2020-Q3
December 2020	2020-2021	2021-Q1 to 2021-Q4	Interest rates	2020-Q3
December 2020	2020-2021	2021-Q1 to 2021-Q4	Other	2020-Q2

APPENDIX C

Table with the variables used in the final forecast.

Source	Variable	Description	Time period	Used in forecast
OECD	Real GDP Growth rate of Portugal	Percentage change from period to period, expressed annually; adjusted for calendar and seasonal factors; chain-linked 2016	1962-Q3 to 2022-Q2	Yes
OECD	Real GDP Growth rate of France	Percentage change from period to period, expressed annually; adjusted for calendar and seasonal factors; chain-linked 2014	1962-Q3 to 2022-Q2	Yes
OECD	Real GDP Growth rate of the United Kingdom	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2018	1962-Q3 to 2022-Q2	Yes
OECD	Real GDP Growth rate of the United States of America	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2017	1991-Q2 to 2022-Q2	Yes
OECD	Real GDP Growth rate of Spain	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2015	1962-Q3 to 2022-Q2	Yes
FRED	Currency Conversions: US Dollar Exchange Rate: Spot, End of Period: USD: National Currency for Portugal	Percentage change from period to period; not adjusted for calendar and seasonal factors	1962-Q3 to 2022-Q2	Yes
OECD	Final consumption expenditure	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1998-Q1 to 2022-Q2	Yes
Engineered	Recession	By "recession" we mean if, in the quarter being analysed, there has been consecutive negative real GDP growth in the previous and current quarter (https://www.imf.org/external/pubs/ft/fandd/basics/recess.htm)	1962-Q3 to 2022-Q2	Yes
Engineered	Year	The year being analysed	1962-Q3 to 2022-Q2	Yes

APPENDIX D

Table with the variables removed from the final forecast after feature selection.

Source	Variable	Description	Time period	Used in forecast
OECD	Real GDP Growth rate of Germany	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2015	1960-Q2 to 2022-Q2	No
OECD	Exports of goods and services growth of Portugal	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1962-Q3 to 2022-Q2	No
OECD	Imports of goods and services growth of Portugal	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1962-Q3 to 2022-Q2	No
OECD	Balance of primary incomes: gross / National income	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1999-Q1 to 2022-Q2	No
OECD	Balance of primary incomes: net / National income	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1999-Q1 to 2022-Q2	No
OECD	Statistical discrepancy (production approach)	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1998-Q1 to 2022-Q2	No
OECD	Compensation of employees	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1998-Q1 to 2022-Q2	No
OECD	Operating surplus and mixed income, gross	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1998-Q1 to 2022-Q2	No
OECD	Domestic demand	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1996-Q1 to 2022-Q2	No
OECD	Individual consumption expenditure: general government	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1998-Q1 to 2022-Q2	No
OECD	Collective consumption expenditure: general government	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1998-Q1 to 2022-Q2	No
OECD	Actual individual consumption	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1998-Q1 to 2022-Q2	No
OECD	Statistical discrepancy (expenditure approach)	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016	1998-Q1 to 2022-Q2	No
FRED	Crude Oil Prices: Brent - Europe	Percentage change from period to period; not adjusted for calendar and seasonal factors	1987-Q4 to 2022-Q2	No

FRED	Business Tendency Surveys (Manufacturing): Capacity Utilisation: Rate of Capacity Utilisation: National Indicator for Portugal	Percentage change from period to period; not adjusted for calendar and seasonal factors	1977-Q2 to 2022-Q2	No
FRED	Business Tendency Surveys (Manufacturing): Orders Inflow: Tendency: National Indicator for Portugal	Percentage change from period to period; not adjusted for calendar and seasonal factors	1987-Q1 to 2022-Q2	No
FRED	World Uncertainty Index: Global: GDP Weighted Average	Percentage change from period to period; not adjusted for calendar and seasonal factors	1990-Q2 to 2022-Q2	No
FRED	Consumer Opinion Surveys: Confidence Indicators: Composite Indicators: OECD Indicator for Portugal	Percentage change from period to period; not adjusted for calendar and seasonal factors	1986-Q3 to 2022-Q2	No
FRED	Consumer Opinion Surveys: Confidence Indicators: Composite Indicators: National Indicator for Portugal	Percentage change from period to period; not adjusted for calendar and seasonal factors	1986-Q4 to 2022-Q2	No
FRED	Interest Rates: Long-Term Government Bond Yields: 10-Year: Main (Including Benchmark) for Portugal	Percentage change from period to period; not adjusted for calendar and seasonal factors	1993-Q4 to 2022-Q2	No
FRED	Interest Rates: 3-Month or 90-Day Rates and Yields: Interbank Rates: Total for Portugal	Percentage change from period to period; not adjusted for calendar and seasonal factors	1986-Q1 to 2022-Q2	No
FRED	Interest Rates: Immediate Rates (< 24 Hours): Call Money/Interbank Rate: Total for Portugal	Percentage change from period to period; not adjusted for calendar and seasonal factors	1999-Q2 to 2022-Q2	No
Engineered	Employement rate	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Total employment / Working Age Population: Aged 15-64: All Persons for Portugal	1998-Q1 to 2022-Q2	No
Engineered	Unemployement rate	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: (Working Age Population: Aged 15-64: All Persons for Portugal - Total employment) / Total employment	1998-Q1 to 2022-Q2	No

Engineered	Value added: agriculture, forestry and fishing percentage	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Value added: agriculture, forestry and fishing / Value added	1998-Q1 to 2022-Q2	No
Engineered	Value added: industry (except construction) percentage	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Value added: industry (except construction) / Value added	1998-Q1 to 2022-Q2	No
Engineered	Value added: construction percentage	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Value added: construction / Value added	1998-Q1 to 2022-Q2	No
Engineered	Value added: services percentage	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Value added: services / Value added	1998-Q1 to 2022-Q2	No
Engineered	Gross fixed capital formation percentage	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Gross fixed capital formation (GFCF)/ Gross capital formation	1998-Q1 to 2022-Q2	No
Engineered	Gross capital formation: changes in inventories and acquisitions less disposals of valuables percentage	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Gross capital formation: changes in inventories and acquisitions less disposals of valuables / Gross capital formation	1998-Q1 to 2022-Q2	No
Engineered	Final consumption expenditure: Households and non-profit institutions serving households change	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Final consumption expenditure: Households and non-profit institutions serving households / Final consumption expenditure	1998-Q1 to 2022-Q2	No
Engineered	Final consumption expenditure: General government change	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Final consumption expenditure: General government / Final consumption expenditure	1998-Q1 to 2022-Q2	No

Engineered	Final consumption expenditure of residents and non-residents on the territory change	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Final consumption expenditure of residents and non-residents on the territory / Final consumption expenditure	1998-Q1 to 2022-Q2	No
Engineered	Final consumption expenditure of durable goods change	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Final consumption expenditure of durable goods / Final consumption expenditure	1998-Q1 to 2022-Q2	No
Engineered	Final consumption expenditure of semi-durable goods, non-durable goods, and services change	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Final consumption expenditure of semi-durable goods, non-durable goods, and services / Final consumption expenditure	1998-Q1 to 2022-Q2	No
Engineered	Taxes less subsidies on products: taxes and subsidies on products net percentage	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: (Taxes less subsidies on products: taxes on products - Taxes less subsidies on products: subsidies on products) / (Taxes less subsidies on products)	1998-Q1 to 2022-Q2	No
Engineered	Harmonized Index of Consumer Prices: Actual Rentals for Housing for Portugal percentage growth	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Harmonized Index of Consumer Prices: Actual Rentals for Housing for Portugal / Harmonized Index of Consumer Prices: All Items for Portugal	1996-Q2 to 2022-Q2	No
Engineered	Harmonized Index of Consumer Prices: Electricity, Gas, and Other Fuels for Portugal percentage growth	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: Harmonized Index of Consumer Prices: Electricity, Gas, and Other Fuels for Portugal / Harmonized Index of Consumer Prices: All Items for Portugal	1996-Q2 to 2022-Q2	No

Engineered	Taxes on production and imports less subsidies: taxes and subsidies on production and imports net percentage	Percentage change from period to period; adjusted for calendar and seasonal factors; chain-linked 2016 Calculation of: (Taxes on production and imports less subsidies: taxes on production and imports - Taxes on production and imports less subsidies: subsidies on production and imports) / Taxes on production and imports less subsidies	1998-Q1 to 2022-Q2	No
Engineered	Quarter	The quarter being analysed	1960-Q2 to 2022-Q2	No
Engineered	Month	The month being analysed	1960-Q2 to 2022-Q2	No

APPENDIX E

Table with the details of the models developed considering each step-ahead.

Step-ahead	Model	Time period	Rolling Window method	Variables	Parameters
1st	GBR	1962-Q3 to 2019-Q2	Anchored	[Real GDP growth of Portugal, Real GDP growth of France, Real GDP growth of Spain, Real GDP growth of United States of America, Real GDP growth of United Kingdom, Currency spot exchange rate of the US dollar to the national currency of Portugal, Recession, Year]	estimators: 150 learning_rate: 0.1 max_depth: 3 min_samples_split: 3 min_samples_leaf: 3
2nd	GBR	1998-Q1 to 2019-Q2	Non-anchored	[Real GDP growth of Portugal]	estimators: 50 learning_rate: 0.05 max_depth: 1 min_samples_split: 2 min_samples_leaf: 6
3rd	GBR	1962-Q3 to 2019-Q2	Non-anchored	[Real GDP growth of Portugal, Real GDP growth of France, Real GDP growth of Spain, Real GDP growth of United States of America, Real GDP growth of United Kingdom, Currency spot exchange rate of the US dollar to the national currency of Portugal, Recession, Year]	estimators: 50 learning_rate: 0.2 max_depth: 1 min_samples_split: 3 min_samples_leaf: 6
4th	GBR	1962-Q3 to 2019-Q2	Anchored	[Real GDP growth of Portugal, Real GDP growth of France, Real GDP growth of Spain, Real GDP growth of United States of America, Real GDP growth of United Kingdom, Currency spot exchange rate of the US dollar to the national currency of Portugal, Recession, Year]	estimators: 50 learning_rate: 0.05 max_depth: 5 min_samples_split: 4 min_samples_leaf: 6

5th	GBR	1962-Q3 to 2019-Q2	Non-anchored	[Real GDP growth of Portugal, Real GDP growth of Spain, Real GDP growth of United States of America, Currency spot exchange rate of the US dollar to the national currency of Portugal, Recession, Year, Date]	estimators: 150 learning_rate: 0.3 max_depth: 3 min_samples_split: 3 min_samples_leaf: 3
6th	GBR	1998-Q1 to 2019-Q2	Non-anchored	[Real GDP growth of Portugal, Final Consumption expenditure of Portugal, Real GDP growth of Spain, Real GDP growth of United States of America]	estimators: 100 learning_rate: 0.1 max_depth: 3 min_samples_split: 2 min_samples_leaf: 1

APPENDIX F

Table with the sensitive analysis of the time periods of 1962, 1979, 1998, 2002, and 2009 for each step-ahead utilizing the same models and parameters as in the final forecast.

Step-ahead	Time period	RMSE	MPE	R^2
1st	1962-Q3 to 2019-Q2	0.6024	-6.70%	0.2925
	1979-Q1 to 2019-Q2	0.5364	-1196.64%	0.3594
	1998-Q1 to 2019-Q2	0.6017	328.51%	-4.4709
	2002-Q1 to 2019-Q2	0.4491	333.08%	-2.2401
	2009-Q1 to 2019-Q2	0.2513	71.62%	-2.8569
2nd	1962-Q3 to 2019-Q2	0.741	-1871.51%	-0.0705
	1979-Q1 to 2019-Q2	0.5463	-2630.83%	0.3355
	1998-Q1 to 2019-Q2	0.3121	-54.9%	-0.4715
	2002-Q1 to 2019-Q2	0.4395	220.1858%	-2.1016
	2009-Q1 to 2019-Q2	0.1899	260.33%	-1.2019
3rd	1962-Q3 to 2019-Q2	0.7573	14.2%	-0.1182
	1979-Q1 to 2019-Q2	0.6697	-1157.53%	0.6975
	1998-Q1 to 2019-Q2	0.6975	183.06%	-6.3522
	2002-Q1 to 2019-Q2	0.7735	-43.17%	-8.6078
	2009-Q1 to 2019-Q2	0.3637	-125.2%	-7.0804
4th	1962-Q3 to 2019-Q2	0.8534	-8.24%	-0.4206
	1979-Q1 to 2019-Q2	0.7732	-673.07%	-0.3311
	1998-Q1 to 2019-Q2	0.7421	178.26%	-7.3226
	2002-Q1 to 2019-Q2	0.8424	234.64%	-10.397
	2009-Q1 to 2019-Q2	0.1232	-141.15%	0.0729
5th	1962-Q3 to 2019-Q2	0.8018	11.59%	-0.2536
	1979-Q1 to 2019-Q2	0.9297	-1225.96%	-0.9244
	1998-Q1 to 2019-Q2	1.2379	1074.24%	-22.154
	2002-Q1 to 2019-Q2	0.7725	177.58%	-8.5816
	2009-Q1 to 2019-Q2	0.5383	604.34%	-16.6966
6th	1962-Q3 to 2019-Q2	0.9544	340.63%	-0.7763
	1979-Q1 to 2019-Q2	0.7210	-1152.94%	-0.3272
	1998-Q1 to 2019-Q2	0.8093	-30,00%	-0.2771
	2002-Q1 to 2019-Q2	0.7250	679.96%	-8.5831
	2009-Q1 to 2019-Q2	0.5660	612.86%	-18.5687

APPENDIX G

Method used to calculate the best model considering the metrics RMSE, MPE, and R^2.

Given the different scales and complexities of the metrics utilized, we employed a system to normalize them by using an arbitrary benchmark value: an RMSE of 0.6618, an MPE of 73.33%, and an R^2 of -0.1959, which were the best values from the original optimal forecast from earlier model development. If a model matches these values, it scores 1; below 1 is better, and above 1 is worse. Although not perfect due to the arbitrary values, this method helps derive sensible conclusions.

APPENDIX H

Spearman's correlation between Portugal's GDP growth and the variables used in the models.

Variable	Correlation
Final Consumption expenditure of Portugal	0.7065
Real GDP growth of France	0.5748
Real GDP growth of Spain	0.5451
Recession	-0.5421
Year	-0.5114
Real GDP growth of the United States of America	0.2174
Real GDP growth of the United Kingdom	0.1406
Currency spot exchange rate of the US dollar to the national currency of Portugal	0.0663

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa