Decision Making and Artificial Intelligence Trends

Volume 1, Issue 1, 34-51.

Predicting U.S. Economic Recessions and Prosperity: a Comparative Study of Machine Learning Models

https://doi.org/10.22034/dmait.2024.199485

Hamidreza Ghadiri[1]*, Mahmoud Gharehgozlou[2]

1. Department of Industrial Engineering and Management Systems, Amirkabir University of Technology, Tehran, Iran.
   *Corresponding Author: hamidreza.ghadiri@aut.ac.ir
2. Department of Industrial Engineering and Management Systems, Amirkabir University of Technology, Tehran, Iran.

ARTICLE INFO

Article history:
Received: 2024/05/15
Revised: 2024/06/12
Accepted: 2024/06/23
Published: 2024/08/27

Keywords:
Economic cycle, Forecasting, Machine learning models, Feature selection

ABSTRACT

The economic cycles, whether it is experiencing recessions or prosperity, serve as the foundation for numerous political and economic decisions. As a result, predicting economic cycles both domestically and globally presents a significant challenge for investors and economic stakeholders. In this study, we investigate the effectiveness of various machine learning (ML) models in predicting economic cycles. Firstly, employs two feature selection methods, including mutual information (MI) and analysis of variance (ANOVA), to select important features. Subsequently, classification models such as Gaussian naïve Bayes, logistic regression, support vector machine (SVM), decision tree, multi-layer perceptron (MLP) neural network, Random Forest (RF), AdaBoost, and voting are utilized to predict economic cycles across various timeframes, ranging from one season to four seasons. This study uses data from the United States (U.S.) economy to evaluate the performance of these models. The results demonstrate the superiority of the ANOVA method in feature selection and the high accuracy of Gaussian naïve Bayes, SVM, and voting models in predicting economic cycles, reaching up to 93% accuracy.

Highlights

- Significant features were selected using both Mutual Information and ANOVA, with ANOVA demonstrating superior performance.
- Economic trends were predicted for different durations, including one season to four seasons.
- Diverse ML models were employed to forecast the U.S. economic trends.
- Logistic regression and SVM exhibited the highest accuracy levels in predicting U.S. economic trends.

1. Introduction

In recent years, there has been a growing interest among political figures, economic stakeholders, and researchers in identifying economic cycles (including recession and prosperity) and their turning points. This focus stems from the profound impact that significant changes in economic activity levels can have on various aspects of society and the economy. Economic downturns, such as recessions, can lead to a cascade of negative effects, including job losses, business closures, and decreased consumer spending. For instance, during the coronavirus recession triggered by the COVID-19 pandemic, financial instability rippled through the U.S. economy and global markets (Albulescu, 2021). The "Great Recession" of 2007-2009, caused by the housing bubble burst, resulted in soaring unemployment rates and a significant decline in the U.S. gross domestic product (GDP) (Axelrad, Sabbath, & Hawkins, 2018). Statistics further emphasize the significance of predicting economic cycles. For instance, in the second quarter of 2022, the U.S. experienced a 0.9% decline in GDP for two consecutive quarters, meeting the technical definition of a recession. While official declarations by bodies like the National Bureau of Economic Research may lag behind real-time indicators, indicators like unemployment rates can signal economic downturns. The ability to predict recessions accurately can help individuals and organizations navigate economic challenges effectively and safeguard their financial interests.

Predicting economic cycles is of paramount importance due to the far-reaching implications on various sectors of society and the economy. The ability to forecast economic recession and prosperity is crucial for policymakers to implement timely monetary and financial measures that can mitigate shocks and stabilize the economy (Gogas et al., 2015). Investors also rely on accurate predictions to adjust their portfolios effectively in economic cycles (Singvejsakul et al., 2019). Notably, recessions can have a profound impact on millions of individuals, affecting their jobs, savings, and overall financial security. Consequently, understanding the warning signs of a recession is essential for individuals and institutions to prepare and minimize the adverse effects on investments and financial well-being. There is a strong emphasis on developing systems capable of providing early warnings about future economic cycles to facilitate informed decision-making (Vrontos et al., 2021). This includes the use of advanced statistical and ML models to analyze historical data and identify patterns, as well as the integration of alternative data sources to enhance the accuracy of forecasts. By leveraging these tools and methods, stakeholders can better understand the dynamics of economic cycles and make more informed decisions to navigate the ever-changing economic landscape.

Previous studies have dedicated significant efforts to crafting predictive models for economic cycles (Hansen, 2024; Borio, Drehmann, & Xia, 2020; Pierdzioch & Gupta, 2020). However, given the diverse nature of recessions and prosperities, influenced by a multitude of factors such as geopolitical events, technological advancements, and global market dynamics, there remains a persistent drive to enhance models' accuracy. The complexity of economic cycles underscores the need for sophisticated and adaptable forecasting tools that can capture the intricacies of these cycles. As a result, numerous research institutions, academic scholars, and industry experts are actively engaged in advancing these forecasting models. Their collaborative efforts aim to refine existing methodologies, integrate cutting-edge technologies like artificial intelligence and ML, and incorporate real-time data sources to improve the precision and reliability of economic cycle prediction (Vrontos et al. 2021).

In the initial stages of this study, the primary focus lay on statistical models (Kauppi & Saikkonen, 2008). However, recent advancements in technology have brought ML models to the forefront, garnering significant attention. Moreover, past research findings suggest that ML models are poised to outperform traditional models in this context (Gogas et al., 2015). The key advantage of ML models over traditional statistical models lies in their ability to handle a larger number of independent features as input, thereby enhancing complexity and accuracy in most scenarios (Jordan & Mitchell, 2015). This aspect holds particular significance as numerous economic variables possess predictive capabilities regarding future economic cycles. Nevertheless, not all features carry equal weight, and a high volume of features can complicate the

model. To address this challenge, employing feature selection algorithms can help mitigate complexity to a certain extent.

The primary aim of this study is to assess various ML models for predicting the economic cycles of the U.S. To accomplish this aim, the effectiveness of eight models - Gaussian naïve Bayes, logistic regression, SVM, decision tree, MLP neural network, RF, AdaBoost, and voting - has been evaluated using 25 input features. Moreover, the performance of these models in forecasting over four different time horizons - three months, six months, nine months, and one year - has been analyzed. To streamline the models' complexity, two feature selection methods- MI and ANOVA - have been implemented.

The main motivation of this paper is to use ML models fed by 25 input features to increase the accuracy of economic cycle forecasting and address the existing research gaps in economic cycle forecasting. In summary, the main research gaps are as follows:

- Insufficient research has been conducted in the realm of predicting economic cycles, with only a limited number of papers, such as Gogas et al. (2015), Vrontos et al. (2021), Borio et al. (2020), and Pierdzioch and Gupta (2020), available in the area.
- Recent research (Vrontos et al., 2021), often focuses on a limited set of features, overlooking the potential benefits of incorporating a broader array of variables to enhance the accuracy of economic cycle prediction.

To address the aforementioned gaps, in the current study, an attempt to enhance economic cycle predictability using more input features is presented.

The remainder of the paper is structured into five sections. Section 2 offers an extensive overview of prior research on predicting economic trends. Section 3 outlines the specifics of the ML models employed in this study. Section 4 undertakes a comparative evaluation of the predictive outcomes derived from these models. Lastly, in Section 5, conclusions are derived from the research outcomes, and potential avenues for future research are proposed, taking into account the constraints of the current study.

2. Literature Review

Considerable efforts have been dedicated to the development of early warning systems over many years. An important study conducted by Burns and Mitchell in 1946 aimed to investigate recurring patterns, referred to as cycles, in a series of time-related data. The main objective of their research was to determine the specific starting and ending points of these cycles (Burns & Mitchell, 1946). Similarly, another significant study initiated by Stock and Watson in 1989 focused on identifying crucial moments, known as turning points, within business cycles. Their research aimed to pinpoint the precise junctures at which overall economic activity shifted from expansion to contraction or vice versa (Stock & Watson, 1989; Stock & Watson, 1992).

The primary goal of these studies was to gain a comprehensive understanding of the patterns and dynamics inherent in economic and financial data. By uncovering these cyclical patterns and identifying turning points, researchers aimed to establish an early warning system that could effectively anticipate and signal changes in economic trends. By analyzing previous research, it becomes evident that there are generally two main approaches to predicting business cycles. In some studies, the focus lies on forecasting a specific economic indicator, typically GDP, for future periods. In contrast, other research adopts a different approach by defining a binary variable and utilizing various models to predict whether upcoming periods will exhibit stagnation or prosperity.

The initial set of research commonly utilizes hidden Markov models for forecasting economic cycles, whereas the subsequent set depends on logit-probit models. For instance, Canova (1994) and Estrella and Mishkin (1998) examined the effectiveness of financial variables in forecasting the economic cycle using a probit model. Other notable research utilizing probit models includes the work of Chauvet and Potter (2005) and Kauppi and Saikkonen (2008). Important conclusions have been drawn from earlier studies indicating the predictive capacity of both macroeconomic indicators and financial variables, notably exemplified by the

stock market index, in forecasting economic cycles. Furthermore, past investigations have demonstrated the efficacy of employing Markov models in this context. Notably, Chauvet and Hamilton (2006) proposed a Markov model in which the recession index and GDP growth rates were configured as components of a Markov chain.

In recent years, as ML models gained popularity, researchers explored their use in economic forecasting. For example, in 2015, Gogas et al. used an SVM with U.S. yield curve data to predict whether GDP would exceed or fall below its trend, achieving significant results (Gogas et al., 2015). Another study conducted by Berge (2015) explored various models for forecasting business cycle turning points. The findings underscored the efficacy of ML models in this endeavor, revealing their significance. The study also found that financial and economic variables were appropriate for a range of prediction horizons. Nyman and Ormerod (2017) carried out a study in 2016 where they employed the RF model to forecast GDP growth rates for one, three, and six future seasons. The findings indicated superior performance of the RF model compared to benchmark models.

James et al. (2019) employed the SVM model to identify the onset and conclusion of the American recession, utilizing National Bureau of Economic Research (NBER) data to pinpoint shifts in economic cycles. Their findings underscored the superior performance of the SVM model. Additionally, they applied the model results to optimize the ratio of stocks to fixed-income bonds in investment portfolios, yielding higher returns and Sharp ratios compared to portfolios with equal allocations. A notable recent study conducted by Vrontos et al. (2021) employed various ML models to forecast the U.S. economy's recession across short-term, medium-term, and long-term horizons. The researchers incorporated a diverse array of financial and economic variables as model inputs. The results revealed that ML models outperformed logic-probit models.

Previous research has explored the application of various ML models, including Bayesian models, tree-based models, and ensemble learning techniques, for economic forecasting. Additionally, some studies have incorporated feature selection methods. However, there is a notable gap in the literature regarding the exploration of combinations of ML models and feature selection methods in this domain. To address this gap, this study aims to develop and compare different ML models using diverse financial and economic variables, to predict the recession or non-recession of the U.S. economy over periods of three, six, nine, and twelve months. Subsequently, the performance of various models will be evaluated and compared to identify the most effective approaches for determining the U.S. economy status across different time periods.

3. Methodology

The methodology of this study is structured to five sequential steps as illustrated in Figure 1. The initial stages involve data collection and feature computation, followed by the application of two distinct feature selection methods to identify important features. Subsequently, ML models are trained using the selected input features and the training dataset. The evaluation process then entails testing these models on a separate dataset. Finally, the predictive performance of various models is compared based on diverse evaluation metrics.

Figure 1. Implementation steps for predicting economic trends using ML

3.1. *Data*

In our study, we used 25 financial and economic variables from the U.S. economy, collected seasonally from the Federal Reserve's website. The selection of the 25 features for our model was a deliberate process, guided by a thorough review of existing literature and empirical studies on economic recession prediction.

We aimed to include a comprehensive set of variables that have demonstrated predictive power in previous research, ensuring our model's robustness and accuracy (Gogas et al., 2015; Vrontos et al., 2021; Nyman & Ormerod, 2017; James et al., 2019). For each variable, we included data from the past four seasons. This gave us a total of 100 features to work with. Our dataset covers the period from January 10, 1993, to January 1, 2022, and includes a binary target variable indicating recession (1) or prosperity (0). We obtained recession and prosperity data from the NBER Institute, identifying 11 recession seasons within our study period. To provide a clearer understanding of our dataset, we have added snippets of our data in Table A1, Appendix A.

The variables we used come from various sectors of the economy and financial markets (Table 1). To predict future seasons, we had 114 records for one season ahead, 113 for two seasons ahead, 112 for three seasons ahead, and 111 for one year ahead. We split the data into training and testing sets, with a 70 to 30 ratio. Our dataset contained no missing values, and we standardized all data to have a mean of zero and a standard deviation of one to account for differences in value ranges.

Table 1: Financial and Economic Variables Used (Source: Federal Reserve)

Code	Feature
1	GDP Growth Rate
2	Personal Consumption Expenditures
3	Trade Balance
4	Personal Saving Rate
5	Unemployment Rate
6	PPI (Total Consumer Goods)
7	Industrial Production
8	Consumer Price Index (CPI)
9	Gross Fixed Capital Formation
10	Total Construction Spending
11	Manufacturers' New Orders
12	Manufacturing Employment
13	Retail Sales
14	New Privately-Owned Housing Units Started
15	Average Sales Price of Houses Sold
16	Continued Claims
17	Total Business Inventories
18	Capacity Utilization
19	Volatility Index (VIX)
20	Term spread - 10-year Treasury yield - 2-year Treasury yield
21	Term spread - 5-year Treasury yield - 2-year Treasury yield
22	Term spread - 2-year Treasury yield - 1-year Treasury yield
23	Gold Price
24	S&P 500 Index
25	Brent Oil Price

To develop models for predicting the recession of the next season, we initially standardized the data to ensure uniform scaling across all variables. Subsequently, the data were split into two groups: a training set comprising 70% of the data and a testing set comprising the remaining 30%, allowing for accurate evaluation of the models' performance. Moving forward, we executed this process for each feature selection algorithm under consideration. A critical aspect of the feature selection discussion is the number of selected features,

as it significantly influences the outcome. In our research, we determined this value through a trial-and-error approach. Our methodology involved selecting several features that maximized the accuracy achievable by the models. Furthermore, given the disparity in the number of recession periods compared to boom periods, we adopted a strategy to balance the classes and improve model training. To achieve this, we applied a weighting Equation (Equation 1) to each class, ensuring that the models were trained more effectively. This approach enabled us to account for the imbalance in the dataset and enhance the predictive capabilities of the models.

$$W_i = \frac{n_0 + n_1}{2 * n_i} \quad (1)$$

In Equation 1 numerator of the fraction is the total number of training data. After the feature selection process and the preparation of the input dataset, the process of training, testing, and calculating the introduced criteria have been done. Hyperparameters have been optimized for SVM, Decision tree, RF, MLP, and AdaBoost models. Of course, because the Bayesian optimizer algorithm does not provide the best possible combination with certainty, the results of the optimization of hyperparameters were compared with the defaults of the models, and those with better-calculated criteria were considered as hyperparameters of the model. The basic criterion of the Bayesian optimizer algorithm was the F1-score because this criterion is more comprehensive than the other metrics, providing a better balance among them. To build the Voting algorithm, four models that had better criteria than others were considered as input to this algorithm at each stage. To reduce variability in results, the Decision tree, RF, MLP, and AdaBoost models were each executed 30 times. The algorithm's performance assessment was made reliable by considering their average results. Finally, for each feature selection algorithm and each prediction phase in the mentioned process, the four criteria were calculated. Since wrongly predicting a recession period as a boom is more consequential than wrongly predicting a boom period as a recession, the most critical evaluation and comparison criterion is Recall. Undoubtedly, a model will be deemed acceptable if it not only fulfills the Recall criteria but also achieves satisfactory results for other relevant criteria.

3.2. *Feature selection*

Feature selection is a critical step in the modeling process, aimed at identifying the most impactful inputs that contribute to effectively predicting the target variable. By focusing on important features, the complexity of the problem is reduced, leading to more interpretable models and improved computational efficiency. Moreover, selecting important features enhances the model's predictive capability by reducing noise and irrelevant information, thus mitigating the risk of overfitting, where the model performs well on the training data but fails to generalize to unseen data.

Various feature selection methods are available in the literature, broadly categorized into filter methods, wrapper methods, and embedded methods. For instance, in cases where the input variables are continuous and the target variable is binary, as in many economic forecasting tasks, methods such as MI and ANOVA are commonly utilized. These methods quantify the relationships between variables and assess the significance of different input features in explaining the variability in the target variable.

In this study, where the input variables are continuous and the target variable is binary, two specific algorithms, MI and ANOVA, have been employed for feature selection. We selected ANOVA and MI for feature selection due to their simplicity, interpretability, and computational efficiency. ANOVA aids in understanding the linear separability of features with respect to the target variable, while MI captures more complex, non-linear dependencies, providing a balanced view of feature relevance. Given the relatively large number of features (25) and the multiple models evaluated, ANOVA and MI offer computationally efficient means of feature selection, avoiding the extensive computational resources required by wrapper methods. Additionally, these methods complement each other by capturing different aspects of feature relevance: ANOVA focuses on linear relationships, and MI identifies non-linear dependencies, ensuring a comprehensive feature selection process.

MI quantifies the relationship between variables by measuring the amount of information obtained about one variable through the other. On the other hand, ANOVA assesses the variability between groups and within groups to determine the significance of different input variables in explaining the variation in the target variable. By utilizing MI and ANOVA methods, this research aims to identify the most relevant features for predicting the economic status of the U.S. economy. This method not only streamlines the input space but also enhances the predictive accuracy of the models, contributing to more robust and effective economic forecasts.

3.2.1. *Mutual Information*

This method assesses the MI between two random variables, providing a measure of their statistical dependence. When two variables are independent, the MI is zero, indicating no relationship between them. Conversely, higher MI values signify a stronger dependence between the variables. This function estimates the output entropy using non-parametric methods, which means it does not rely on specific assumptions about the underlying data distribution.

MI serves as a powerful tool in feature selection by quantifying the amount of information shared between input and output variables. By analyzing the MI between each input variable and the target variable, researchers can identify the most relevant features for predicting the target outcome. This process helps streamline the modeling process by focusing on key variables that contribute significantly to the predictive performance of the model. Furthermore, the non-parametric nature of this method makes it versatile and applicable to various types of data distributions. Unlike parametric methods, which make assumptions about the data's underlying distribution, non-parametric approaches provide robust results without imposing such constraints. As a result, MI-based feature selection offers a flexible and reliable means of identifying important variables in predictive modeling tasks, contributing to more accurate and interpretable models.

3.2.2. *ANOVA*

ANOVA, or analysis of variance, is a parametric statistical test employed to assess whether the means of two or more data samples, typically three or more, are derived from the same underlying distribution. Essentially, ANOVA serves as a specialized form of the F test in statistics, which calculates the ratio between variance values. The F statistic, or F test, represents a class of statistical tests that quantify the ratio between different variance values. This can include comparisons between the variances of two distinct samples or the variances explained and unexplained by a statistical model, such as ANOVA. In the context of ANOVA, the method specifically evaluates the F statistic, often referred to as the ANOVA F test.

ANOVA is particularly useful when dealing with numerical input variables and a categorical target variable. By analyzing the variation within and between different groups or categories, ANOVA helps ascertain whether there are statistically significant differences in the means of the groups being compared. This method plays a pivotal role in hypothesis testing and model building, enabling researchers to determine the impact of various factors on the outcome of interest. Moreover, ANOVA's parametric nature ensures robustness and reliability when analyzing data that adheres to certain distributional assumptions.

3.3. *Classification models*

Classification algorithms are a fundamental component of ML, particularly in scenarios where the target variable is unpredictable or random. These algorithms typically fall into two main categories: individual models and ensemble models. Individual models, such as logistic regression, SVM, decision trees, and k-nearest neighbors, operate independently to make predictions based on the input data. They have been selected for this study based on their historical performance, especially in the field of forecasting economic cycles.

In addition to individual models, ensemble models have been employed to explore whether combining different algorithms can improve predictive accuracy. Ensemble models, including techniques like RF, AdaBoost, and gradient boosting machines, leverage the collective wisdom of multiple models to achieve superior performance compared to individual models alone.

Overall, by leveraging both individual models and ensemble models, this research aims to comprehensively investigate and improve upon existing methods for forecasting economic trends.

3.3.1. *Individual models*

3.3.1.1 Gaussian naive Bayes

The Bayesian classification technique is commonly utilized as a straightforward method for categorizing and labeling data. It encompasses a family of algorithms that operate under the assumption of feature or variable independence. Maximizing the likelihood function is central to most approaches and simple Bayesian models. Despite its simplicity and straightforward assumptions, the simple Bayes classification technique is effective in addressing real-world problems. The cornerstone of all algorithms in this category is conditional probability.

One prominent algorithm in this category is Gaussian Naïve Bayes, which assumes a normal distribution for the data distribution within each class. To construct this model, one approach involves assuming Gaussian distribution for the data and independence among variables. Subsequently, the distance of each data point from the mean of each class is computed and normalized by the standard deviation of the class. This process facilitates the determination of the probability of each data point belonging to each class.

3.3.1.2 Logistic regression

Logistic regression is a classification algorithm utilized to categorize data into distinct classes. Unlike linear regression, where the target variable is continuous, logistic regression deals with target variables that take on two or more discrete values. This model calculates the probability of assigning data to each category. The logistic regression Eq is as follows:

$$f(x) = \frac{e^{b_0+b_1x_1+b_2x_2+\cdots+b_kx_k}}{1+e^{b_0+b_1x_1+b_2x_2+\cdots+b_kx_k}} \quad (2)$$

In logistic regression, model parameters are determined by maximizing the likelihood function. If the output probability from the logistic function is below 0.5, the model predicts class zero; conversely, if the probability is above 0.5, class one is predicted.

3.3.1.3 Support vector machines

The SVM is a powerful ML algorithm used for both classification and regression tasks. SVM aims to find an optimal hyperplane that separates data points of different classes with the maximum margin, thereby enhancing the algorithm's ability to generalize to unseen data. It achieves this by transforming the data into a higher-dimensional space using kernel functions, allowing for nonlinear decision boundaries. SVM seeks to minimize classification errors while maximizing the distance between the decision boundary and the closest data points, known as support vectors. This characteristic makes SVM robust to outliers and capable of handling datasets with complex structures. SVM has been widely applied in various domains, including image classification, text categorization, and bioinformatics, due to its versatility and strong theoretical foundation.

3.3.1.4 Decision tree

A decision tree is an ML method used to organize an algorithm. It comprises nodes forming the root tree, with the root being the initial node without incoming edges. Internal nodes, or test nodes, have outgoing edges, while the rest are leaves. Test nodes divide the sample space into subspaces based on a discrete function of input values. Typically, each test node partitions the sample space based on a single feature's value, or for numeric features, within a specified range. Each leaf is assigned a class representing the most

suitable target value or a probability vector indicating the likelihood of target attribute values. Samples are classified by traversing from the root to the leaf, guided by test results. Parameters like tree depth can be adjusted in decision tree algorithms to mitigate overfitting or excessive complexity.

3.3.1.5 Multi-layer perceptron

The research conducted by Qi (2001) has effectively underscored the advantageous applications of Neural Networks (NNs) within the realm of economic cycle analysis. MLP represents a neural network architecture featuring numerous layers of interconnected neurons. Employing feedforward propagation, it processes data and introduces nonlinearities via activation functions. Trained through supervised learning, utilizing backpropagation to refine weights, MLP demonstrates versatility in modeling intricate data relationships. Widely applied in classification and regression tasks, MLPs necessitate meticulous tuning to avert overfitting and maximize effectiveness.

3.3.2. *Ensemble models*

3.3.2.1 Random Forest

The RF is an ensemble learning technique that operates by aggregating the predictions of multiple decision trees. During training, a specified number of decision trees are constructed, each trained on a random subset of the training data and a random subset of the features. This randomness introduces diversity among the trees, making them less correlated and less prone to overfitting. During prediction, each decision tree independently outputs a class label (for classification) or a numerical value (for regression). The final prediction of the RF is then determined by aggregating the individual predictions through a voting mechanism (for classification) or averaging (for regression).

The key principle behind RF is the wisdom of crowds: by combining the predictions of multiple weak learners (individual decision trees), RF leverages the collective knowledge of the ensemble to make more accurate and robust predictions than any single tree alone. Random Forests are highly effective for handling high-dimensional data with complex interactions, and their ability to capture non-linear relationships makes them suitable for a wide range of tasks.

3.3.2.2 Adaptive Boosting

AdaBoost is an ensemble learning technique used for classification and regression tasks. It works by combining the predictions of multiple weak learners, typically decision trees, to create a strong classifier or regressor. The key idea behind AdaBoost is to sequentially train a series of weak learners, with each subsequent learner focusing on the instances that were misclassified by the previous ones. This adaptive learning process assigns higher weights to misclassified instances, effectively prioritizing them in subsequent training iterations. During training, AdaBoost iteratively updates the weights of training instances based on their classification accuracy, focusing more on misclassified instances in each subsequent iteration. This emphasis on difficult-to-classify instances allows AdaBoost to gradually improve its performance over multiple rounds of training. During prediction, AdaBoost combines the predictions of all weak learners through a weighted sum, where the weight of each weak learner depends on its classification accuracy. The final prediction is determined by a majority vote (for classification) or weighted average (for regression) of the individual weak learners.

The iterative and adaptive nature of AdaBoost allows it to construct a strong model that performs well on the entire dataset, leveraging the collective knowledge of the ensemble to make more accurate predictions than any single weak learner alone. Additionally, AdaBoost is robust to overfitting and can handle noisy data effectively.

3.3.2.3 Voting classifier

The voting classifier is an ensemble learning model. It operates by combining the outcomes of multiple learning algorithms, selecting the output with the highest number of votes as the final prediction, typically resulting in enhanced accuracy compared to individual learning methods. There are two primary approaches to employing this algorithm: firstly, through voting for the predicted classes of each learning algorithm, with

the most frequently occurring class chosen as the output. Secondly, by evaluating the probability of data belonging to each class, which entails calculating the average probabilities if all individual algorithms can compute them. Subsequently, the class with the highest average probability is identified as the output. This approach offers a robust mechanism for making predictions, leveraging the collective knowledge of diverse learning algorithms to achieve superior performance.

3.4. *Hyperparameters Selection*

In many ML models, choosing the right hyperparameters greatly impacts the accuracy of the final model. Thus, in this study, we employed the Bayesian optimization algorithm to select these hyperparameters. Bayesian optimization uses Bayes' theorem to guide the search for the best hyperparameters, especially useful for complex and costly functions to evaluate. This algorithm constructs a probabilistic model of the target function, called the surrogate function, and utilizes an acquisition function to choose which hyperparameters to evaluate next. By considering past evaluations, it efficiently explores the hyperparameter space, typically requiring fewer attempts to find the best solution compared to other methods. This makes Bayesian optimization an effective tool for optimizing ML models.

3.5. *Evaluation criteria*

The implementation results were compared with four criteria: accuracy, precision, recall, and F1-score. Table 2 shows how to calculate these criteria.

Table 2: Criteria used to compare models

Criteria	Formula
Accuracy	$\frac{TP + TN}{TP\,TN\,FP\,FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-score	$\frac{2 * Precision * Recall}{Precision + Recall}$

4. Empirical results

The empirical results are presented in this section. Table 3 shows the number of features selected by each feature selection algorithm for each prediction time period. As can be seen, the number of selected features varies depending on the model and the prediction time period. Tables 4-10 show the out-of-sample performance of the underlying models for each forecasting horizon.

Table 3: Number of Selected Features for Each Time Horizon

Forecasting Horizon	Algorithm	Number of Features Selected
One Season Ahead	MI	20
One Season Ahead	ANOVA	20
Two Seasons Ahead	MI	30
Two Seasons Ahead	ANOVA	30
Three Seasons Ahead	MI	28

Three Seasons Ahead	ANOVA	28
Four Seasons Ahead	MI	30
Four Seasons Ahead	ANOVA	30

Table 4: Calculated metrics for one-season-ahead prediction models with the ANOVA method

ANOVA	Accuracy	Recall	Precision	f1-score
Gaussian naive bayes	0.942857143	0	0	0
Logistic regression	0.971428571	1	0.666666667	0.8
SVM	0.971428571	1	0.666666667	0.8
Decision Tree	0.543809524	0.583333333	0.061468058	0.106001742
RF	0.705714286	0.8	0.224729682	0.307767791
MLP	0.557142857	1	0.117387335	0.209577358
AdaBoost	0.485714286	1	0.1	0.181818182
Voting	0.971428571	1	0.666666667	0.8

The performance of ML models on the one-season-ahead prediction, trained using the ANOVA feature selection algorithm, is summarized in Table 4. From the results, it is evident that the Logistic regression, SVM, and Voting models achieved the highest accuracy (97.14%). However, the Logistic regression model exhibited the lowest F1 score (0.8), while the SVM and Voting models demonstrated the highest F1 score (0.8). On the other hand, the Decision tree model yielded the lowest accuracy (54.38%) and F1 score (0.106), indicating its unsuitability for this particular task. The RF model exhibited moderate accuracy (70.57%) and F1-score (0.308), suggesting that it is a viable option for this task. Conversely, the MLP and Ada Boost models displayed low accuracy (55.71% and 48.57%, respectively) and F1-score (0.209 and 0.182, respectively), indicating their unsuitability for this task.

Table 5: Calculated metrics for one-season-ahead prediction models with the MI method.

MI	Accuracy	Recall	Precision	f1-score
Gaussian naive bayes	0.942857143	0	0	0
Logistic regression	0.771428571	1	0.2	0.333333333
SVM	0.971428571	0.5	1	0.666666667
Decision Tree	0.712380952	0.433333333	0.098888889	0.145295691
RF	0.911428571	0.383333333	0.275	0.317777778
MLP	0.748571429	0.95	0.182415362	0.303755243
AdaBoost	0.485714286	1	0.1	0.181818182
Voting	0.942857143	0	0	0

Table 5 presents the results of applying the MI feature selection algorithm to train various ML models for the one-season-ahead prediction task. The results indicate that the SVM model achieved the highest accuracy (97.14%). The Logistic regression and Voting models followed closely with the second-highest accuracy (94.29%). However, it should be noted that the Logistic regression model exhibited the lowest F1-score (0), while the SVM and Voting models demonstrated the highest F1-score (0.667). The Decision Tree model displayed moderate accuracy (71.24%) and F1-score (0.145), suggesting that it holds potential for this task. Further tuning may be necessary to enhance its performance. In contrast, the RF model exhibited high accuracy (91.14%) and a relatively high F1-score (0.318), indicating its suitability for this task. Similarly, the MLP model demonstrated moderate accuracy (74.86%) and F1-score (0.304).

The ANOVA feature selection algorithm consistently yielded better accuracy across the models. The Logistic regression + ANOVA, SVM + ANOVA, and Voting + ANOVA models emerged as the top performers, considering all four metrics. These three models achieved remarkable accuracy, successfully predicting all recessions (two recession periods) in the subsequent period with a Recall score of 1. Only one prosperity period was misclassified as a recession during the testing phase, showcasing the models' exceptional forecasting capabilities.

However, it is important to note that the remaining models, despite undergoing hyperparameter optimization, displayed high recall but comparatively lower scores in other evaluation metrics. This suggests that forecasting economic conditions proves to be a challenging task for these methods. Nevertheless, based on the impressive results obtained by the Logistic regression and SVM models, it is evident that reliable models can be developed to provide early warnings of recessions. Moreover, an interesting observation is that the Voting algorithm did not significantly outperform the best individual learning algorithm. This finding implies that combining multiple algorithms does not necessarily lead to improved results in this particular context.

Table 6: Calculated metrics for two-seasons-ahead prediction models with the ANOVA method.

ANOVA	Accuracy	Recall	Precision	f1-score
Gaussian naive bayes	0.941176471	0	0	0
Logistic regression	0.852941176	1	0.285714286	0.444444444
SVM	0.823529412	1	0.25	0.4
Decision Tree	0.507843137	0.6	0.052277622	0.09377018
RF	0.367647059	1	0.087044499	0.159886304
MLP	0.785294118	1	0.217705628	0.356926037
AdaBoost	0.305882353	1	0.083529412	0.153535354
Voting	0.852941176	1	0.285714286	0.444444444

Table 6 presents the results of ML models on the two-seasons-ahead prediction, where the models were trained using the ANOVA feature selection algorithm. The Gaussian Naïve Bayes model attained the highest accuracy of 94.11%, followed by the logistic regression and voting algorithm with an accuracy of 85.29%. Notably, both the Logistic regression model and voting algorithm achieved the highest F1-score of 0.444, while the MLP model secured the third highest F1-score of 0.35. Additionally, the logistic regression and voting algorithm demonstrated the highest precision of 0.2857.

Table 7: Calculated metrics for two-seasons-ahead prediction models with the MI method.

MI	Accuracy	Recall	Precision	f1-score
Gaussian naive bayes	0.941176471	0	0	0
Logistic regression	0.705882353	1	0.166666667	0.285714286
SVM	0.823529412	1	0.25	0.4

Decision Tree	0.473529412	0.566666667	0.043492523	0.079323204
RF	0.903921569	0.466666667	0.210634921	0.278624339
MLP	0.823529412	1	0.25	0.4
AdaBoost	0.058823529	1	0.058823529	0.111111111
Voting	0.823529412	1	0.25	0.4

Table 7 presents the results of ML models trained using the MI feature selection algorithm for the Two-Seasons-Ahead Prediction task. The Gaussian Naïve Bayes model achieved the highest accuracy of 94.11%, followed by the RF model. Both the SVM and Voting models, as well as the MLP model, attained the highest precision and F1 score, with values of 0.25 and 0.4, respectively. The results of predicting one season ahead versus two seasons ahead show a comparable pattern, although there's a noticeable decline in overall model accuracy, which is understandable given the extended forecasting horizon. However, what's particularly noteworthy is the superior models' ability to accurately anticipate all recessionary periods. This indicates that the reduced forecast accuracy is primarily attributable to misidentifying periods of prosperity as a recession. Tables 8-9 show the performance of different ML models on the three-seasons-ahead prediction.

Table 8: Calculated metrics for three-seasons-ahead prediction models with the ANOVA method.

ANOVA	**Accuracy**	**Recall**	**Precision**	**f1-score**
Gaussian naive bayes	0.941176471	0	0	0
Logistic regression	0.794117647	1	0.222222222	0.363636364
SVM	0.941176471	0	0	0
Decision Tree	0.535294118	0.566666667	0.045645339	0.083306184
RF	0.923529412	0.65	0.472222222	0.494126984
MLP	0.337254902	1	0.083062073	0.153186756
AdaBoost	0.058823529	1	0.058823529	0.111111111
Voting	0.941176471	0.5	0.5	0.5

Table 9: Calculated metrics for three-seasons-ahead prediction models with the MI method.

MI	**Accuracy**	**Recall**	**Precision**	**f1-score**
Gaussian naive bayes	0.941176471	0	0	0
Logistic regression	0.529411765	1	0.111111111	0.2
SVM	0.941176471	1	0.5	0.666666667
Decision Tree	0.571568627	0.666666667	0.109309189	0.163356268
RF	0.848039216	1	0.288917749	0.445614756
MLP	0.401960784	1	0.092483279	0.168908127
AdaBoost	0.058823529	1	0.058823529	0.111111111

Voting	0.735294118	1	0.181818182	0.307692308

As depicted in Table 8, the Gaussian naive Bayes, SVM, and Voting models exhibited the highest accuracy (94.12%) when employing ANOVA feature selection. Among these, the voting algorithm demonstrated the highest precision and f1-score, registering at 0.5 and 0.66, respectively. When employing MI feature selection, the Gaussian naive Bayes and SVM models showcased the highest accuracy (94.12%) in Table 9. Among these, the SVM model displayed the highest precision and f1-score, achieving 0.5 and 0.66, respectively. It's worth mentioning that in this section, while most models successfully forecasted all recessions, the aforementioned models exhibited relatively fewer errors in accurately predicting periods of prosperity. Tables 10-11 show the performance of different ML models on the four-seasons-ahead prediction.

Table 10: Calculated metrics for four-seasons-ahead prediction models with the ANOVA method.

ANOVA	Accuracy	Recall	Precision	f1-score
Gaussian naive bayes	0.941176471	0	0	0
Logistic regression	0.764705882	1	0.2	0.333333
SVM	0.911764706	0.5	0.333333333	0.4
Decision Tree	0.488235294	0.666666667	0.072156456	0.123112
RF	0.900980392	0.4	0.226111111	0.28619
MLP	0.719607843	0.916666667	0.17057383	0.281917
AdaBoost	0.058823529	1	0.058823529	0.111111
Voting	0.911764706	0.5	0.333333333	0.4

Table 11: Calculated Metrics for Four-Seasons-Ahead Prediction Models with the MI Feature Selection Algorithm

MI	Accuracy	Recall	Precision	f1-score
Gaussian naive bayes	0.941176471	0	0	0
Logistic regression	0.705882353	1	0.166666667	0.285714286
SVM	0.676470588	1	0.153846154	0.266666667
Decision Tree	0.52745098	0.533333333	0.056363212	0.095079365
RF	0.921568627	0.35	0.327222222	0.304047619
MLP	0.442156863	1	0.09721498	0.176923224
AdaBoost	0.058823529	1	0.058823529	0.111111111
Voting	0.764705882	1	0.2	0.333333333

When forecasting the recession for the upcoming year, the ANOVA feature selection algorithm has shown relatively superior performance overall. Notably, Logistic regression + ANOVA achieved the best results in terms of recall, while SVM + ANOVA stood out across other evaluation criteria. Despite a general decrease in forecasting accuracy over a one-year period, it's noteworthy that all recessionary periods were still successfully predicted.

5. Conclusion

The analysis of economic cycles has long been a focal point for investors, economic stakeholders, and policymakers. Recent advancements in technology, particularly in the realm of ML models, have spurred research into predicting economic cycles. Some studies have explored the use of feature selection algorithms to streamline the problem space and enhance forecasting accuracy. However, there remains a gap in the literature concerning a comprehensive comparison of methods and the integration of feature selection methods with ML for forecasting economic cycles. This study aims to address this gap by predicting economic cycles across four time periods using a dataset comprising 25 financial and economic features. By leveraging feature selection methods, specifically focusing on seasonality variations ranging from one to four seasons, this research endeavors to enhance the accuracy of economic cycle predictions in the U.S.

Achieving a high accuracy rate in predicting economic cycles holds substantial significance due to the profound impact economic cycles have on various sectors of society and the economy. The high accuracy rate enhances the reliability of the predictions, which in turn has several practical applications. Accurate predictions of economic cycles can be instrumental for policymakers, enabling them to implement timely and effective measures to mitigate economic shocks and stabilize the economy. For instance, early identification of an impending recession allows for proactive monetary and fiscal policies to cushion the negative effects on employment, income, and overall economic stability.

For investors and financial institutions, a high accuracy rate in predicting economic cycles can significantly improve investment strategies and risk management. Accurate predictions enable investors to adjust their portfolios in anticipation of economic downturns or upswings, optimizing returns and minimizing losses. This can also aid financial institutions in better managing credit risks and making informed lending decisions based on anticipated economic conditions. Furthermore, businesses can benefit from these predictions by making informed operational and strategic decisions. For example, during anticipated economic downturns, businesses might choose to scale back investments, reduce inventories, or delay expansion plans to conserve resources. Conversely, during periods of predicted economic growth, businesses may increase investments, expand operations, and hire more employees to capitalize on favorable economic conditions.

Accurate predictions can also help individuals make informed decisions regarding their personal finances, such as adjusting savings and spending behaviors, managing debt, and planning for major life events like buying a home or retirement. By having a reliable forecast of economic conditions, individuals can better safeguard their financial well-being and navigate economic uncertainties.

Moreover, the ability to predict economic cycles with high accuracy is crucial for various sectors beyond just finance and business. For example, government agencies and non-profits can better plan and allocate resources for social programs, public services, and community support initiatives during different phases of the economic cycle. This ensures that support systems are strengthened during downturns when they are most needed.

In this study, two feature selection methods, ANOVA and MI, were employed to identify significant features for prediction. Subsequently, gaussian naïve Bayes, logistic regression, SVM, decision tree, multilayer perceptron, random forest, AdaBoost, and voting models were utilized to forecast periods of recession and prosperity based on the selected important features. The findings of this study indicated that the ANOVA method outperformed the MI method. Furthermore, gaussian naïve Bayes, SVM, and voting models demonstrated superior accuracy compared to other models across all periods, successfully predict recessionary periods with precision. However, despite the high prediction accuracy achieved by the other models, they did not show success in terms of other performance evaluation metrics. Furthermore, the improved accuracy in predicting recessionary periods for the next three seasons compared to the subsequent two seasons implies a delayed influence of specific features on the economy. As a result, in line with previous research, both Gaussian naïve Bayes and SVM models exhibited consistent accuracy

throughout all examined time frames, highlighting their potential for developing an early warning system. Overall, based on the comprehensive analysis of the results obtained, the following findings can be drawn:

- The ANOVA algorithm consistently outperformed the mutual algorithm in feature selection across most cases, establishing it as the superior choice for this task.
- The Voting algorithm failed to enhance accuracy compared to the best-performing individual learning algorithms across all prediction periods, suggesting its limited utility.
- Logistic regression and SVM consistently demonstrated the highest accuracy among all learning algorithms, albeit with minor variations across different prediction periods.
- Variations in the selected features for each prediction period suggest that certain variables are more suitable for short-term forecasting, while others are more effective for long-term predictions.
- An important observation is the ability of top models to accurately predict recession periods. However, these models occasionally misclassified some prosperity periods as recession, which, while a low-cost error in practical forecasting, underscores their high operational value. Predicting recessions accurately is crucial for investors and policymakers, making this achieved accuracy highly desirable.
- The generally low accuracy of most algorithms, except for Logistic regression and SVM, implies that modeling this problem effectively with all algorithms is unfeasible. Only the two mentioned models consistently deliver satisfactory results.

To enhance the precision of economic cycle forecasting models, future study endeavors could explore the utilization of more sophisticated ML models and assess their outcomes. Additionally, incorporating additional financial and economic features, particularly the yield curve, could further enrich the predictive capabilities. In this study, a binary classification approach was employed for economic cycle prediction; expanding the classification scheme could yield a more nuanced evaluation of economic trends. Furthermore, to validate the robustness of the findings, applying the developed models to different countries and comparing the outcomes with those of this study could provide valuable insights into model generalizability and cross-country economic trend analysis.

References

Albulescu, C. T. (2021). COVID-19 and the United States financial markets' volatility, Finance Research Letters, 38, 101699. https://doi.org/10.1016/j.frl.2020.101699

Axelrad, H., Sabbath, E. L., Hawkins, S. S. (2018). The 2008–2009 Great Recession and employment outcomes among older workers, European Journal of Ageing, 15, 35-45. https://doi.org/10.1007/s10433-017-0429-0

Berge, T. J. (2015). Predicting recessions with leading indicators: Model averaging and selection over the business cycle, Journal of Forecasting, 34(6), 455-471. https://doi.org/10.1002/for.2345

Borio, C., Drehmann, M., Xia, F. D. (2020). Forecasting recessions: The importance of the financial cycle, Journal of Macroeconomics, 66, 103258. https://doi.org/10.1016/j.jmacro.2020.103258

Burns, A. F., Mitchell, W. C. (1946). Measuring business cycles, National Bureau of Economic Research.

Canova, F. (1994). Were financial crises predictable? Journal of Money, Credit and Banking, 26(1), 102-124.

Chauvet, M., Hamilton, J. D. (2006). Dating business cycle turning points. Contributions to Economic Analysis, 276, 1-54. https://doi.org/10.1016/S0573-8555(05)76001-6

Chauvet, M., Potter, S. (2005). Forecasting recessions using the yield curve, Journal of Forecasting, 24(2), 77-103. https://doi.org/10.1002/for.932

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

Gogas, P., Papadimitriou, T., Matthaiou, M., Chrysanthidou, E. (2015). Yield curve and recession forecasting in a machine learning framework, Computational Economics, 45, 635-645. https://doi.org/10.1007/s10614-014-9432-0

Hansen, A. L. (2024). Predicting recessions using VIX–yield curve cycles, International Journal of Forecasting, 40(1), 409-422. https://doi.org/10.1016/j.ijforecast.2023.04.002

James, A., Abu-Mostafa, Y. S., Qiao, X. (2019). Machine learning for recession prediction and dynamic asset allocation, The Journal of Financial Data Science, 1(3), 41-56. https://doi.org/10.3905/jfds.2019.1.007

Jordan, M. I., Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Kauppi, H., Saikkonen, P. (2008). Predicting US recessions with dynamic binary response models, The Review of Economics and Statistics, 90(4), 777-791. https://doi.org/10.1162/rest.90.4.777

Lahiri, K., Moore, G. H. (Eds.). (1991). Leading economic indicators: New approaches and forecasting records, Cambridge University Press.

Liu, W., Moench, E. (2016). What predicts US recessions? International Journal of Forecasting, 32(4), 1138-1150. https://doi.org/10.1016/j.ijforecast.2016.02.007

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

Pierdzioch, C., Gupta, R. (2020). Uncertainty and forecasts of US recessions, Studies in Nonlinear Dynamics & Econometrics, 24(4), 20180083. https://doi.org/10.1515/snde-2018-0083

Qi, M. (2001). Predicting US recessions with leading indicators via neural network models, International Journal of Forecasting, 17(3), 383-401. https://doi.org/10.1016/S0169-2070(01)00092-9

Singvejsakul, J., Chaiboonsri, C., Sriboonchitta, S. (2019). The dependence structure and portfolio optimization in economic cycles: An application in ASEAN stock market, In Integrated Uncertainty in Knowledge Modelling and Decision Making: 7th International Symposium, IUKM 2019, Nara, Japan, March 27–29, 2019, Proceedings 7 (pp. 161-171).

Stock, J. H., Watson, M. W. (1989). New indexes of coincident and leading economic indicators, NBER Macroeconomics Annual, 4, 351-394.

Stock, J. H., Watson, M. W. (1992). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience.

Vrontos, S. D., Galakis, J., Vrontos, I. D. (2021). Modeling and predicting US recessions using machine learning techniques, International Journal of Forecasting, 37(2), 647-671. https://doi.org/10.1016/j.ijforecast.2020.08.005

Appendix A. Data Snippets

Table A1. Example Records from the Dataset

date	GDP Growth Rate(0)	GDP Growth Rate(1)	GDP Growth Rate(2)	Personal Consumption Expenditures (1)	Personal Consumption Expenditures (2)	...	class
1993-10-01	0.019127888	0.011	0	4487.189	4418.581	...	0
1994-01-01	0.014530625	0.019	0	4552.651	4487.189	...	0
1994-04-01	0.018449328	0.015	0	4621.223	4552.651	...	0
1994-07-01	0.011610984	0.018	0	4683.163	4621.223	...	0
1994-10-01	0.016943354	0.012	0	4752.761	4683.163	...	0
1995-01-01	0.008987044	0.017	0	4826.713	4752.761	...	0
1995-04-01	0.007804539	0.009	0	4862.436	4826.713	...	0
1995-07-01	0.013471579	0.008	0	4933.609	4862.436	...	0
1995-10-01	0.01164383	0.013	0	4998.662	4933.609	...	0
1996-01-01	0.01233592	0.012	0	5055.655	4998.662	...	0
1996-04-01	0.020889962	0.012	0	5130.615	5055.655	...	0
1996-07-01	0.012270629	0.021	0	5220.499	5130.615	...	0
1996-10-01	0.015786073	0.012	0	5274.505	5220.499	...	0
1997-01-01	0.012456035	0.016	0	5352.763	5274.505	...	0
1997-04-01	0.018674691	0.012	0	5433.105	5352.763	...	0
1997-07-01	0.016903505	0.019	0	5471.267	5433.105	...	0
1997-10-01	0.011899585	0.017	0	5579.179	5471.267	...	0