Recession Prediction: Final Report

Garrett Laverty, Ethan Chapel, Samuel Asebrook

I. Abstract

This report presents a comprehensive study on the development of our machine learning model to accurately predict upcoming recessions over large time frames. We are assessing multiple variables of proven relevance with the aim of improving upon existing models. Our approach is simple: create models based on previous work and our collected data that has proven to have strong correlations to U.S recessions, and then combine their predictions to create a more accurate overall model.

Recessions can have severe impacts on individuals and families, often leaving them unprepared and in difficult circumstances. Early warnings of such economic downturns can be crucial in allowing people to prepare. Traditional economic indicators, such as GDP growth rates, unemployment figures, and consumer spending patterns have served as primary metrics for determining potential recessions. However, these indicators often exhibit lagging behaviors, providing limited foresight, not to mention that they can be difficult to interpret at times. The integration of advanced machine learning techniques offers a promising avenue to enhance predictive accuracy and timeliness in identifying recessionary trends.

As previously stated, current models have demonstrated high accuracy, however they lack consistency across different decades. This raises questions about their future utility, particularly when we are inevitably faced with changes in data that is already volatile. In addition, the most accurate models rely on analyzing multiple months of data to make a single prediction. Thus, our goal was to create a model capable of consistent predictions across long periods of time utilizing data from only a single month.

Our models featured consistently high accuracy and recall across the board, regardless of the time period in which the evaluation was being performed. This was significant given the variability of economic data as a whole, let alone the variability that has been expressed through existing models. Our combined model featured a near 93% accuracy and 90% recall, which is excellent and speaks to the efficacy of our methods.

Although he have not achieved accuracy as high as that of existing models, we are only 3%-4% off while our model only requires one month of data to make a prediction. This indicates that our model has practical uses when data is limited and the multi-month changes between variables cannot be evaluated.

II. Introduction

We are assessing the variables that signal recessions in an attempt to create a machine learning algorithm that can accurately predict upcoming recessions over large time frames. Countless models have been built to tackle this problem, but we believe there is potential to improve upon the previous work in terms of consistency between decades. To do this, we are going to create models based upon previous work, and attempt to find a way of combining these models predictions to create a more accurate overall model. Each model will be utilizing data that has already been collected and proven to have strong correlations to U.S recessions.

Recessions are very detrimental to a significant number of people when they occur. If not prepared for them, they can leave families in shambles with no roof over their heads and no food on their plates. The best way to combat this is to have money set aside for such events. Many people don't have the extra funds to set aside though, so a warning ahead of time would be helpful to allow people to prepare.

Traditionally, economic indicators, such as GDP growth rates, unemployment figures, and consumer spending patterns, have served as primary metrics for determining potential recessions (see Fig. 1 below). However, these indicators often exhibit lagging behaviors, providing limited foresight into impending recessions. Hence, the integration of advanced machine learning techniques offers a promising avenue to enhance predictive accuracy and timeliness in identifying recessionary trends.

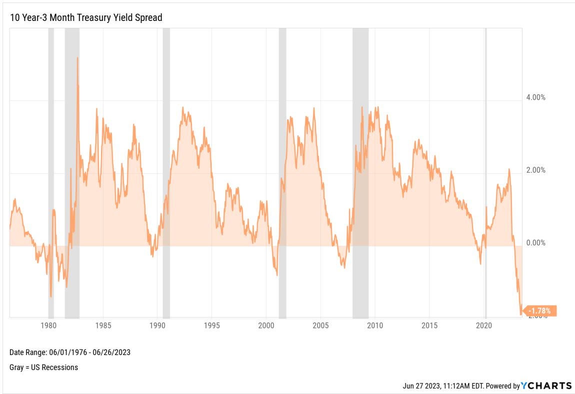

Fig. 1

This figure displays the relationship between the 10-year vs 3-month yield spread (orange) and the occurrence of U.S recessions (gray)[9].

The data that we chose explores two avenues of ideology. First of all, we looked at data for things that experienced significant crashes in recent economic recessions that have hit our country. This includes things such as credit card interest rates, mortgage loan interest rates, etcetera. Second, we looked at varying methods of investment. This includes bond maturity rates, NASDAQ ratings, etcetera. All of this data was sourced from the St. Louis Federal Reserve's online economic data repository (FRED).

Current models have been proven to be capable of very high accuracy (see Fig. 2 for an example). However, none have been able to achieve consistency between decades. As a result, it is reasonable to question whether they will be useful going into the future. Thus, the creation of a model that can adapt to fundamental changes in the data and create accurate predictions between decades will be an immense leap forward.

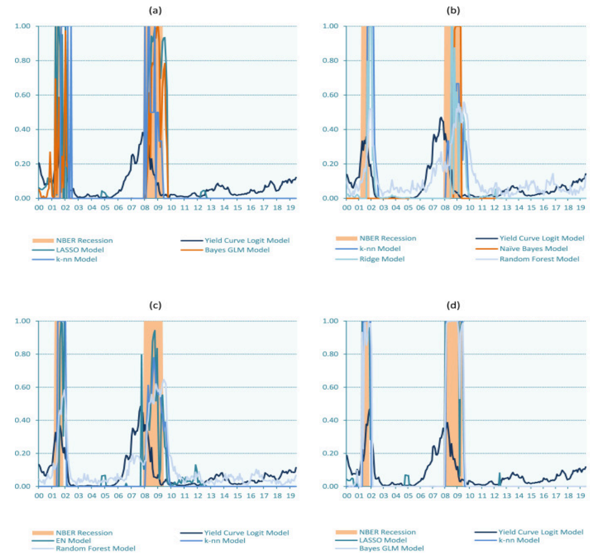

Fig. 2

Predictive capabilities of current recession predicting models [8].

Due to the extremely volatile nature of economics, creating a model capable of consistent predictions across such long periods of time has been proven an immense challenge. In addition, current models which are tailored to specific time frames are quite accurate, with some reaching as high as 97% [8]. Maintaining such accuracy will also prove difficult.

We are aiming to improve prior work by creating a model capable of making accurate predictions across multiple decades. We are putting together a solid data set with many factors that could potentially point towards recessions. We now need to determine which factors have changes that most frequently occur right before a recession occurs. We will then use those to develop individual models which will contribute to an overall prediction.

Addressing these challenges presents an opportunity to make significant contributions to both academic research and practical applications. By harnessing the power of machine learning, we aim to develop a predictive framework capable of offering early warnings for recessions, thereby facilitating proactive decision-making and risk management strategies. Furthermore, the development of interpretable models can enhance our understanding of the underlying drivers of economic downturns, enabling policymakers to formulate targeted interventions and foster economic resilience. Overall, this endeavor holds the potential to advance the frontier of recession forecasting and foster a more resilient and adaptive economic ecosystem.

We will delve into contributions in methodology in later sections of this report. In summary, we have found that the combination of similar existing models into a single prediction with weights based on error metrics prove a useful advancement in improving the consistency of recession prediction models.

III. Background and Related Work

Prior works on this topic are very extensive, with many of the first applications of machine learning to economics beginning in the 1990's. Previous work utilizes a plethora of models and methods for estimation, the most commonly occurring include: support vector regression, random forest, long-short term memory models, binary probit/logit regression, penalized likelihood models, ridge logit, lasso logit, elastic net binary logit, linear and regularized discriminant analysis, classification and regression trees, CART recursive partitioning, bagging, random forests, adaptive and gradient boosting, Naive Bayes learning, KNN, Markov-switching models, and artificial neural networks. Benchmarks have been used as additional means of validation. These benchmarks most often consist of comparisons to single variable predictors, most commonly the 10-year vs 3-month Treasury bill yield curve. There are a number of simple baseline prediction models that are used to compare results to as well, such as the vector autoregression model [9]. As of now, hundreds of variables have proven useful in predicting recessions, but the most consistent appear to be the 10-year vs 3-month Treasury bill yield curve, real money supply growth, real GDP growth, and average hourly earnings [8].

Some models are able to display surprising accuracy, with the most accurate models having 96%-97% accuracy within specific time frames [8]. This predictive accuracy is reinforced when models are combined into a majority vote system [1]. The majority vote system takes predictions from each model and returns the most occurring. However, the current difficulty with predictive models is the ability to maintain that accuracy across a large time frame. The models which reach high levels of accuracy only tend to be able to do so for specific time horizons, often less than a decade, as they overfit their predictions to that period. The models that are currently the best at achieving consistent predictions include mostly penalized likelihood binary models.

While the prior work done on this topic is very impressive, there are still criticisms that can be made. To start, many journals written did not consider nearly enough economic variables to build comprehensive models [3]. There have been exceptions, such as "*Modeling and predicting U.S. recessions using machine learning techniques*" which included 56 variables in their analysis and rated each one based on impact to performance. Many also failed to make proper mention on how their data was processed prior to use, such as whether certain variables were normalized or over what periods of time each variable covered. The only criticism that can be applied to all prior work as a whole is failure to produce models which can maintain accurate predictions across multiple decades. It has been concluded that models which can select the most impactful variables perform best, but creating a system which also takes into account interplay between variables could be just as beneficial.

The main source of prior work that we will be building on is the work done by Spyridon Vrontos, John Galakis, and Loannis Vrontos [8]. These individuals provide a very in-depth analysis of many of the methods which will be discussed in later sections of this report. While we won't be utilizing any existing code, many of our ideas will be originating from, and trying to improve upon this work.

We are aiming to improve prior work by creating a model capable of making accurate predictions across multiple decades. The process for achieving this goal is outlined in our "Methodology" section.

Below is a brief description of each of the methods we used during the course of our project:

- The Linear Regression model takes in a linearly related dependent and independent variable and attempts to model the relationship between the two.

- The SVM model attempts to find a hyperplane that separates the classes provided to the feature space. Maximizing the margin created allows for better performance and robustness against outliers.

- Random Forest builds multiple decision trees and uses averaging and voting in order to make a prediction.

- Gaussian Boosting builds multiple decision trees sequentially where each tree corrects the errors present in the previous tree.

- Randomized Search determines hyperparameters by taking a randomized subset of combinations and evaluating them.

- The Probit model takes in a vector of variables that influence a single dependent variable into being classified into a binary classification. It then uses those variables to determine the probability of the dependent variable being either class.

- The Lasso Logit models perform variable selection, choosing those which are the most strongly correlated with the dependent variable. Coefficients for this process are found by maximizing each log-likelihood function and imposing shrinkage penalties [7].

- The K-nearest neighbors approach uses the differences between variables of our data and variables for the time period we are testing for and finds the most closely related periods, which it will then classify the time period we are testing for to align with the majority of the nearest periods.

- Decision trees work by having a split at each characteristic in the data, creating a bottom layer of all possible combinations of data characteristics. Each new point of data will follow a certain path of the tree down until the end, which it will be classified as the most frequently occuring outcome within that final subset.

- Ridge classification is similar to linear regression and simply aims to minimize variance.

- Regularized discriminant analysis is a solution to unstable covariance matrices found in linear discriminant analysis [8].

- Linear discriminant analysis seeks to create the best linear combination of the predictors.

- Bagging (bootstrap aggregating) improves predictive accuracy by generating multiple samples from the training data and training models for each sample [6]. It then takes the result that occurs the most as the final prediction. This is used to improve accuracy if we use a system like decision trees.

- Adaboost (adaptive boosting) iteratively trains weak classifiers and gives higher weightage to data which is misclassified. This cycle of resampling and training creates increasingly stronger classifiers [8].

IV. Data

The data that we are using is directly sourced from the St. Louis Federal Reserve, more specifically the Federal Reserve Economic Data (FRED) repository [4]. The datasets that we sourced from the FRED are all based on a monthly cycle and include:

- Recession indication
- National unemployment rate
- Average percentage of income Americans are saving
- Single-month inflation representation (average CPI)
- Decade-based real interest rate
- Average interest rate on credit cards
- Decade-based mortgage interest rate
- Ratio of available automobiles to automobiles sold
- Annual automobiles sold without regard to inventory
- Index changes for the NASDAQ100
- 30-year based mortgage interest rate
- 10 year vs 3 month yield curve
- Ratio of securities in banks
- Federal reserve borrowings

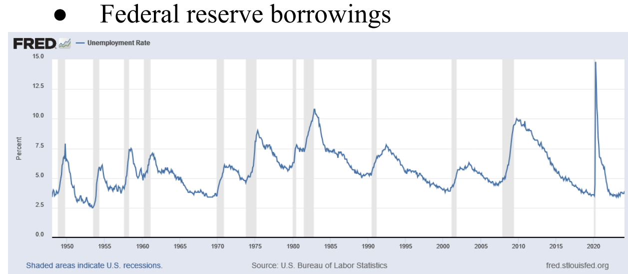

Fig. 3

FRED data for unemployment rates in the United States [4]

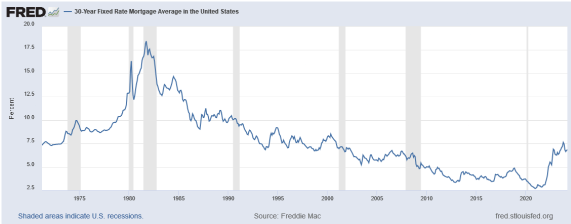

Fig. 4

FRED data for interest rates on fixed 30-year mortgages [4]

The above graph in fig. 3 shows the FRED data for unemployment rates in the United States [4]. Our intentions regarding the use of this data is proving a correlation between unemployment rates at any given time and whether or not the country is actively in a recession.

The same can be said for the FRED data for interest rates on fixed 30-year mortgages as seen in fig. 4. There are periods of trends in every dataset we are utilizing for our models, and our effort is to correlate them to the state of our national economy.

Since most values for these datasets are defined, the only data processing we have done is dropping any entries with a NaN value present. We have also conducted an indicator selection process that is described in full in the experimental results section of this report.

V. Methodology

Our approach includes the individual development of the following models: Linear Regression, SVM, Adaboost, Gaussian Boosting, Probit, Lasso Logit, KNN, Decision Trees, Random Forest, Ridge Classification, and Linear Discriminant Analysis. We have decided to replace our RDA model with an LDA model as our dataset is very unbalanced. This caused the RDA model to always perform equal to or slightly worse than our LDA model, making it unnecessary to include. We will also apply bagging techniques to improve performance where deemed necessary. All the listed models will be fit to 70%-80% of our data. Then, Random Search will be performed to find the weights for each model that produced the highest recall and

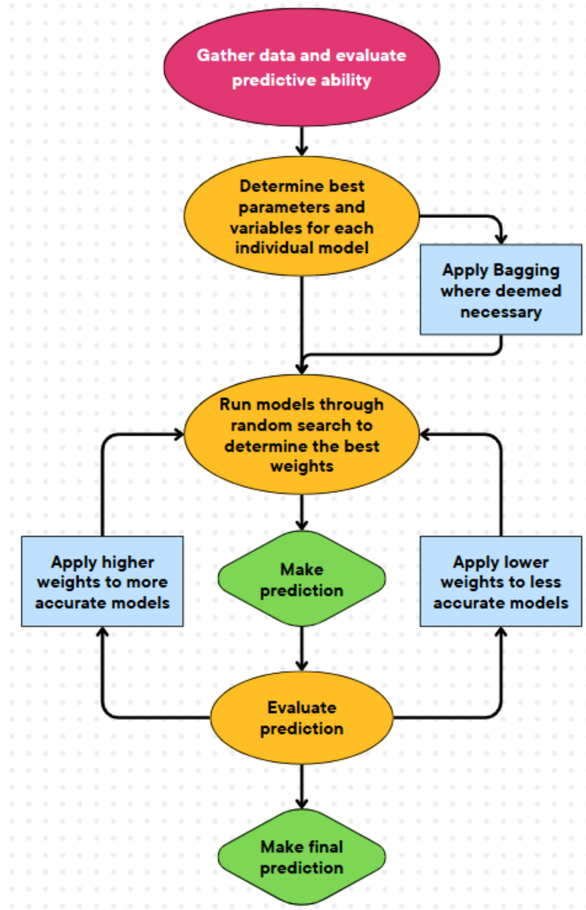

Fig. 5

Schematic detailing our current process for completing our outlined methods.

accuracy, with recall being prioritized. For the testing data, if the final prediction from the Random Search exceeds a given threshold (also found via Random Search) then the final prediction is marked as a recession. This entire process is outlined in Fig. 5 above.

While essentially every possible model has already been applied to the task of predicting recessions, we failed to find a study that combined models in the way we have listed above. While some have used a majority vote system with multiple models, the inclusion of weights specific to each individual model's predictive ability has not been explored in full. Most previous studies attempt to find the most accurate single model. As a result, we believe that combining the predictions from each model in such a manner has the potential to increase accuracy.

VI. Experimental Results

In Figure 6 below, we have compiled a list of our models and the associated evaluation metrics for each model. To reiterate how the final prediction is calculated, each model is trained using our training data, then a list of weights is created with each weight corresponding to an individual model. Then, a random search is performed creating several hundred random combinations of weights and applying them to the models predictions of 0 (not a recession) or 1 (is a recession) for a piece of data. Thus if a model's ideal weight is found to be 1.3 and it predicts a result to be a recession (1.0) the final value of that model will return 1.3. In the Final Predictor, each of these values from all the individual models is added together, and if the total exceeds a given threshold, then the final prediction will be a recession.

In the chart below, the threshold found in the random search algorithm was 1.76, meaning that some models were accurate enough to only require two predictions of recession in the individual models for the final prediction to be a recession. For example, a prediction of recession (1) from the Linear Regression model and the KNN model would be enough by themselves to have the final prediction be that of recession (1) since their combined value would be 2.26 which exceeds the threshold.

	Accuracy	Precision	Recall	Specificity	F1	Mean Squared Error	Weight In Final Prediction
Linear Regression	0.969135	1.0	0.5	1.0	0.667	0.025	1.27
SVM	0.956790	1.0	0.3	1.0	0.462	0.033	0.52
KNN	0.987654	1.0	0.8	1.0	0.889	0.008	0.99
Adaboost	0.987654	1.0	0.8	1.0	0.889	0.016	1.20
Random Forest	0.975309	0.875	0.7	0.9934	0.778	0.041	0.73
Bagging	0.969136	1.0	0.5	1.0	0.667	0.024	1.47
Ridge Classification	0.845679	0.273	0.9	0.8421	0.419	0.172	0.81
Lasso Logit	0.956790	0.8	0.4	0.9934	0.533	0.024	0.93
Linear Discriminant Analysis	0.969136	1.0	0.5	1.0	0.667	0.024	1.22
Probit	0.944444	0.538	0.7	0.9605	0.609	0.049	0.25
Gaussian Boosting	0.956790	0.714	0.5	0.9868	0.588	0.016	0.28
Final Predictor	0.925926	0.45	0.9	0.9276	0.6	0.016	

Fig. 6

Chart of all individual model metrics and final prediction metrics.

Overall, our individual models have seen an improvement in all metrics since the previous report. One reason for the increase in our models accuracy has been the replacement of the certain variables in our dataset (see Fig. 7 below). We are currently excluding the index change of the NASDAQ 100, automobile sales ratio, and credit card interest rates from the data being used with our models and have replaced them with the 10 year bond maturity, 3 month bill market rate, securities in banks, and federal reserve borrowing. This is because the excluded data had too many NaN values limiting our training data as any preprocessing made it difficult for trends to remain detectable by many of our individual predictive models. The new data has proven useful, providing a small boost to metric scores as they have decent predictive capabilities and they don't have any holes in their data. Another reason for the improvement is the changes made to the parameters of the individual models. Each parameter of every model was adjusted until a set of parameters that consistently produced the best results was chosen.

VII. Discussion

Since our last report we have seen a significant improvement in our predictive capabilities. We are now confident that combining the models listed above into an overall prediction with applied weights has a considerable positive impact on the final predictions. While the accuracy of our final model is below most of the individual models, the recall is the highest among them, tying only with Ridge Classification (which has a much lower overall accuracy). This is important to note as the instances of "recession" are far fewer than instances of "not recession", so when we have a datapoint indicating a recession it is crucial that our model predicts that point accurately (with our current setup, recall measures this proportion). In addition, our final prediction is much more consistent than all of the individual models which can have large variations in results depending on our train/test split. Our final recession predicting model consistently has above 90% accuracy and recall, and often has accuracies in the 95%-99% range and recall at 100% depending on the value of the random_state parameter in the train/test split. This indicates that we are on track of accomplishing our goal of creating a model that is consistently accurate across many decades.

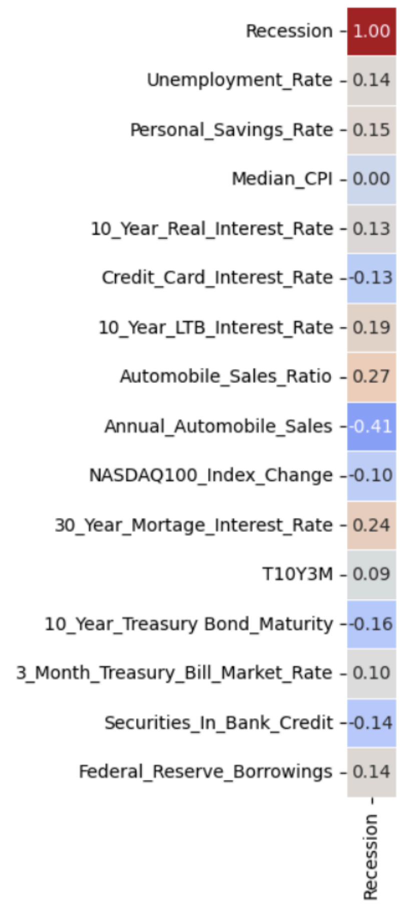

Fig. 7
Individual Indicator Correlations

In comparison with currently existing recession predicting models, ours has slight advantages and disadvantages. At the moment, the baseline of our model's accuracy is lower at about 90% while the baseline for some current models is 97% accuracy. However, most current models use adjacent data in the evaluation, for example, in determining a prediction for March of 2023 the models will look at data for January, February, and March to analyze the differences between periods as well. However, our model is able to have fairly high accuracy scores while only relying on the data for that single month. Another cause of our lower accuracy is our dataset. Current models have many more features to compare, with larger models using upwards of 50-60 different variables. Our model currently uses 12, as we have removed some variables that proved to hinder our accuracy and added additional variables to replace them. We have also conducted more fine tuning on the parameters for each individual model, as there was still room for improvement. We started out on this path through writing Python code to analyze the efficacy of the parameters of our models in regard to the predictive power of every model. We struggled to implement search algorithms, especially given the number of combinations needed to be processed, and the approach that we took was not producing the kinds of results that we were looking for.

To pivot from this, we ended up conducting manual testing of the models with multiple combinations of parameters and making the necessary changes to improve our model's performance. We made many changes, but those which yielded the best results was reducing the number of neighbors in the KNN model from five to three and increasing the number of estimators in our adaboost model from fifteen to thirty. This resulted in a highly effective implementation, though not as optimized as it could have possibly been through the utilization of automated search algorithms.

VIII. Conclusion

In conclusion, after applying the weights to each model, we were able to predict recessions with an accuracy consistently above 90%. Unfortunately, in its current state, our model was unable to outperform the most accurate models currently existing today. However, it has become apparent that the inclusion of a weighted combined model system does have potential to yield better results than a majority vote system. While a minimum of 97% accuracy is a difficult goal to achieve, we have still managed to create a model capable of producing high accuracy predictions while only evaluating one month at a time in comparison to the current models which also evaluate the feature changes in the months leading up to the month being predicted. This fact means our model does have potential applications in instances where data is sparse or unavailable prior to the month being evaluated. However, it also indicates our model may be limited when dealing with variables that only show strong correlations to recessions when evaluated over multiple months and determining the rate of change.

Future research utilizing our methods could benefit largely from two adjustments. The first being the inclusion of more variables as our dataset was limited to those publicly available, making many variables incomplete which would result in worse predictions. A research team capable of retrieving complete datasets from private sources of more accurate variables could create a much more accurate model. Additionally, creating an alternative model that is able to take multiple months into account has the potential to greatly improve accuracy in certain scenarios as they are shown to be more accurate than their single-month counterparts when equipped with large, full datasets. As a result, a system that can switch between our current single-month approach and a new multi-month approach depending on the available data in use could be a significant step forward in recession-predicting machine learning models.

References

[1] A. James, Y. Abu-Mostafa, and X. Qiao, “Nowcasting recessions using the SVM machine learning algorithm,” SSRN Electronic Journal, Jun. 2018. doi:10.2139/ssrn.3316917

[2] A. Estrella and F. S. Mishkin, “Predicting U.S. recessions: Financial variables as leading indicators,” Review of Economics and Statistics, vol. 80, no. 1, pp. 45–61, Feb. 1998. doi:10.1162/003465398557320

[3] E. Clower, “Predicting Recessions with Machine Learning Techniques | Aptech,” Feb. 21, 2023. https://www.aptech.com/blog/predicting-recessions-with-machine-learning-techniques

[4] Federal Reserve Bank of St. Louis, “FRED Economic Data,” *Stlouisfed.org*, 2024. https://fred.stlouisfed.org/

[5] J. H. Friedman, “Regularized discriminant analysis,” Journal of the American Statistical Association, vol. 84, no. 405, p. 165, Mar. 1989. doi:10.2307/2289860

[6] L. Breiman, “Bagging predictors,” Machine Learning, vol. 24, no. 2, pp. 123–140, Aug. 1996. doi:10.1007/bf00058655

[7] R. Tibshirani, “Regression shrinkage and selection via the lasso,” Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267–288, Jan. 1996. doi:10.1111/j.2517-6161.1996.tb02080.x

[8] S. D. Vrontos, J. Galakis, and I. D. Vrontos, “Modeling and predicting U.S. recessions using machine learning techniques,” International Journal of Forecasting, vol. 37, no. 2, pp. 647–671, Apr. 2021, doi: doi.org/10.1016/j.ijforecast.2020.08.005

[9] YCharts, “The inverted yield curve: What it means and how to navigate it,” Nasdaq, https://www.nasdaq.com/articles/the-inverted-yield-curve-what-it-means-and-how-to-navigate-it

[10] Y. Huang and E. S. Yan, “Economic recession forecasts using machine learning models based on the evidence from the COVID-19 pandemic,” Modern Economy, vol. 14, no. 07, pp. 899–922, 2023. doi:10.4236/me.2023.147049