Forecasting of Economic Recession using Machine Learning

Vedanta Bhattacharya
Department of Electronics and Communication Engineering
Amity University Uttar Pradesh
Noida, Uttar Pradesh, India
vedanta1209@gmail.com

Ishaan Srivastava
Department of Electronics and Communication Engineering
Amity University Uttar Pradesh
Noida, Uttar Pradesh, India
im.ishaanlenovo@gmail.com

***Abstract*— A significant, highly widespread, and long-lasting drop in economic activity is described as a recession. Since the time it touches the high end of the previous expansion to the trough, economists all around the world tend to search for the length and the factors in between this length that constitutes the recession. Recessions may only last a few months, but it may take years for the economy to bounce back and reach back to its previous high point.**

In this paper, we investigate various datasets and its pros and cons and try to give the shot to speculate the chances of recession, based on Machine Learning techniques and backed by concrete datasets for better approximation of result. Within a Machine Learning framework, we try to put in the values, specifically the Chicago Federation National Activity Index (CFNAI) and its 2 components: Monthly Average of 3 and CFNAI Diffusion Index from a period of March, 1967 to June, 2022, along with the S&P-500 index and various parameters associated with it (like 10yr Tbond, % Tbond, 2 year spread fedrate, CPI etc) from the day of the recorded period. Using unscaled, scaled and tuned models of machine learning technique, the model was able to predict the chances, ranging from 87% when the models used were unscaled, to slightly more than 94% for scaled models. Both the models are different from each other, and are compiled together for comparison and the purpose of having a surety about the model being correct in these regards.

Keywords*— *Recession, Machine Learning, Regression, ML algorithms

I. INTRODUCTION

Recession is defined as a period of economic downturn wherein the economy starts to contract and it mostly gets evident in the country's growth as the country's GDP plunges and the stock market feels the reverberations and starts to go down, more commonly termed as "bear market". A series of local or international market shocks can trigger a global financial crisis, which can then turn into a worldwide economic crisis because of the interconnection of the financial markets. In other cases, a single economic power that is quite huge and a member of the "big-enough" economies to generate turmoil in other countries could be the source of an economic crisis. This was the case, for instance, with the subprime crisis or what is commonly called as the Great Recession of 2008, which began in the US and spread to other European nations and inevitably worldwide as a problem of sovereign debt. The studies argue that firms' risk management and financing policies had a significant impact on the degree to which firms were impacted by the financial crisis [1] (Brunnermeier, 2009). Exploration of the 2008 crisis by Erkens et all [2] hypothesise that this is due to (a) firms with more independent boards raised more equity financing during the crisis, which caused existing shareholders' wealth to be transferred to debtholders, and (b) firms with higher institutional ownership took on more risk before the crisis, which resulted in greater investor losses during the economic crisis.

Failure to forecast recessions is a recurring theme in economic forecasting. The challenge of predicting output gaps is extremely contemporary and has profound implications for economic policy. Of course, early notice of an impending output decline is crucial for policymakers, who can then quickly alter monetary and fiscal measures to either prevent a recession or lessen its effects on the real economy. The NBER estimates [3] that only five recessions have occurred in the United States since 1980, compared to 34 altogether since 1854. The plummet that occurred due to the double-dip falls in the early 1980s and the global financial crisis of 2008 can be termed as those which were far worse than either the Great Depression or the depression of 1937–1938.

With more potential regressors than there are observations, the usage of Machine Learning models is now proven to be able to lead and supervise massive volumes of data and provide high assurances in terms of accuracy of the eventual model. Chen et al. considered the healthcare as one of the sector affected by the Great Recession of 2008 [4] and thus, examined the paper based on the health care expenditures along the health care spending distribution, based on the Medical Expenditure Panel Survey (MEPS) dataset from 2005-06 and 2007-08. To determine the various relationships between the recession and health care spending along the health care expenditure distribution, quantile multivariate regressions are used.

II. LITERATURE REVIEW

Although machine learning algorithms have long been employed in categorization issues, they are now increasingly being used in the social sciences, notably in the field of

financial sectors (Dabrowski, 2016) [5]. The hidden Markov model, switching linear dynamic system, and Naive Bayes switching linear dynamic system models were all implemented in this work. The hidden Markov Model stated that: First, the assumption of a limited horizon states that the probability of being in a state at a given time depends only on that state (t-1) and second, the assumption of a stationary process states that, given the current state, the conditional (probability) distribution over the next state remains constant.. Nyman and Ormerod used the ML's Random Forest technique for the dataset between 1970(Q2)-1990(Q1), and from 1990-2007 which comprised of GDP growth period [6]. This model was able to predict the result of about six quarters ahead and the results were stronger in case of the economy of the UK compared to the US. In another paper [7] both of them extended their analysis by looking at how each of the explanatory variables affected the Great Recession of the late 2000s. They were able to further the investigation by breaking down business and non-financial household debt into separate categories and discovered that the Great Recession was significantly influenced by both household and non-financial company debt though, their explanatory models exhibit significant non-linearity.

Using the 3 months to 2 year period, the Treasury term spread was used as a benchmark by Liu and Moench [8] and they paid particular attention to the subject of whether or not the leading indicators surveyed before in the literatures go beyond the Treasury term spread to provide insight into potential future recessions. The Italian economy was employed as the database, and machine learning guided tools were used as the analysis method, in the paper [9] presented by Paruchuri, who investigated the idea of machine learning in economic forecasting. By examining the financial characteristics that can be used as a recession indicator, Estrella and Mishkin [10] conducted additional analysis of the US recession, wherein they were making conclusions from the result they got (about one to eight quarters ahead) and Stock prices, currency exchange rates, interest rates, and monetary aggregates were assessed separately as well as in relation to other financial and non-financial indicators.

III. METHODOLOGY AND DATASET

A. Dataset

The first thought of using the dataset was the quarterly GDP growth rate of major economies around the world, but it was quickly rejected since there weren't too many available data (eg- after the disintegration of USSR, the quarterly data for Russia is available only after 1993, while the Chinese maintained the data only after 1990). Yearly data wasn't favourable either due to it being too little (maintained and in the open domain only from the 1960s onwards, with major data from countries like France available only after 1970s), so there wasn't any continuity. The reason why the search for the dataset was diluted to only the US economy is due to the fact that being the largest economy of the world, with major trade governed through the US Dollar, whenever it faces any economic stress, the world faces the aftershock.

One of the major indicators about any recession that has been triggered by the US was the CFNAI (Chicago Fed National Activity Index). This index is a weighted average of 85 actual economic activity indicators for the preceding period. A single, summary measurement of a common element in the entire national economic statistics is offered by the CFNAI. As a result, changes in the CFNAI throughout time closely mirror times of economic expansion and contraction.

The CFNAI Diffusion Index represents the difference between the total absolute value of the weights for the indicators that are hidden underneath whose contribution to the Chicago Fed National Activity Index is always affirmative in a given month and the total absolute value of the weights for those indicators whose contribution is negative or neutral in that same month over a three-month period. The CFNAI Diffusion Index, whenever below the threshold of -0.35, have indicated that the economy is under recession. Fisher et all in their paper outlined the failure of other policies and other Federation banks to capture the accurate representation of recession and thereon, gave some statistical model that was the determinant of the current Index and also calculated the magnitude to which it could go [19].

The other dataset that was used for the paper was the S&P-500 data. The Standards & Poors-500 (hereon S&P-500) is an index very similar to the Nifty-100 or the Nifty-50 back in India, which has been used since ages to measure the stock performances of 500 major corporations that are listed on American stock exchanges, from NYSE to NASDAQ. It is one of the most common and the most sought-after equity index. More than $5.4 trillion was invested in assets linked to the index's performance by the end of the year 2020, with this list reaching its highest point on 2 January, 2022.

One of the key reason to pick it was the historical evidence and the volume of the data that it presents, from December 1927 to current, updated daily (as and when the market trades). The paramaters associated with S&P-500, along with the high and low price and trading volume includes 10yr Tbond, % Change Tbond ,2yr SpreadFedrate ,% Change Fedrate, Nonfarm Payrolls , % Change Payrolls , CPI and % Change CPI Date.

Since the features we needed for our dataset were not conveniently included in a downloadable dataset, we had to download each feature separately and combine them together into one dataframe. We were able to pull each economic feature separately from FRED (Federal Reserve Economic Data from the Federal Reserve Bank of St. Louis) using Quandl, which also had the added bonus of automatically calculating selected transformations if we chose to do so, and the financial feature was downloadable from Yahoo! Finance so we downloaded the dataset and created the transformed variables in Excel and imported the dataset as a CSV. We created Recession labels from a list of start and end dates. We finally concatenated each feature and the labels using an inner join to create one data-frame. After creating a correlation heatmap, we selected the features we wanted to include in our final dataset (each of the features fall under a certain category; employment, monetary policy, inflation, bond market, or stock market). Finally

we saved this dataset as a CSV and performed some descriptive statistics on the data.

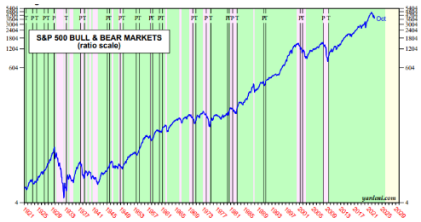

Fig 1: The Ratio Scale of S&P-500 Bull and Bear Market (Credit: Yardemi)

B. Importance of Data Scaling

It is typical for data to contain scales of values that fluctuate from variable to variable. A variable could be expressed in feet, metres, and so forth.

Some machine learning techniques, such as normalisation, which scales all variables to values between 0 and 1, perform significantly better when all variables are scaled within the same range. This has an influence on algorithms like support vector machines and k-nearest neighbours as well as methods like linear models and neural networks that use a weighted sum of the input.

As a result, scaling input data is a recommended practise, as is experimenting with different data transforms, such as employing a power transform to make the data more normal (better suit a Gaussian probability distribution).
This is true for output variables, also known as target variables, such as numerical values expected when modelling regression predictive modelling problems.
It is frequently useful to scale or convert both the input and target variables in regression situations.
It is simple to scale input variables. You can use the scale objects manually in scikit-learn, or the more handy Pipeline, which allows you to chain a series of data transform objects together before utilising your model.
The Pipeline will fit the scale objects to the training data and apply the transform to new data, such as when using a model to predict something.

a. Scaling of Target Variables

There are two methods for scaling target variables.
The first option is to handle the transform manually, while the second is to use a new automatic method.

1. Transform the target variable manually.
2. Transform the target variable automatically.

*1)**When the Target Variable is changed manually***
Effectively managing the scaling of the target variable manually entails manual constructing and the application of the scaling object to the data.

It involves the following steps:

- Create the object that tends to be used for transform e.g. a MinMaxScaler.
- Then, Fit this transform function on the training dataset.
- Apply the transform function to the subdivided dataset consisting of training and testing datasets.
- Invert the transform for any forecast that was made.

2)***Automatic Transform of the Target Variable**:* An approach that could be followed instead of the aforementioned one is to automatically take charge of the transform and inverse transform.This can be achieved by using the "TransformedTargetRegressor" object that tend to encapsulate a given model and a scaling object.The "TransformedTargetRegressor" tends to prepare the transform of the targeted variable with the help of the same training data used to fit the model. On applying that inverse transform on any new data provided when calling the predict() function, it returns those predictions in the correct scale. To use the TransformedTargetRegressor, it is defined by specifying the model and the transform object to use on the target.

b. Rescaling Options

This wasn't covered before, but there are essentially just two options for rescaling features. Consider the case where x1 runs from -1 to 1, whereas x2 ranges from 99 to 101: both of these features have (about) the same standard deviation, but x2 has a much bigger mean. Consider the following scenario: x1 is still between -1 and 1, while x2 is between -100 and 100. They have the same mean this time, but x2 has a significantly greater standard deviation. Gradient descent and similar algorithms can become slower and less trustworthy in each of these circumstances. As a result, we want to make sure that all features have the same mean and standard deviation.

C. Model used in the project

There were two main models that were used in the making of this project, namely

- Linear Regression
- Logistic Regression

Linear Regression:
A linear method that corresponds to the proximity of the relationship between a scalar response and one or more interpretive variables is linear regression (the responses mentioned above are also known as dependent and independent variables). In contrast to multivariate linear regression, which predicts numerous correlated variables dependent on each other, instead of just a single scalar variable, this phrase is more specific.

In linear regression, linear predictive functions are used to model associations, with the parameters that remain unknown in the model being estimated from the data. Such type of models are called linear models. The conditional mean of the response is typically considered to be a related function of the values of the explanatory variables (or predictors); the conditional median or some another quantile is usually placed.

The model parameters are quite easy to be understood due to the linear shape. Additionally, linear model theories (which are mathematically easy and fair) are widely known. Furthermore, a lot of contemporary modelling tools are built on the foundation of linear regression [20]. If we take an example, linear regression more often than not provides a good round-off to the underlying regression function, especially when the sample size is tiny or the signal is very faint.

The first regression analysis method for which the researchers undertook an in-depth research and saw plenty of its use in actual applications was linear regression. This is because models with linear dependence on their uncertain parameters are in a simple way fits the linear models than their non-linear counterparts. Plus, it is simpler to determine the statistical attributes and characters of the resulting estimators.
The equation to show the linear regression is given as:

$$y = mx + c + e$$

Where *m*= the line's slope
c=intercept; and
e= representation of the error that the model may have.

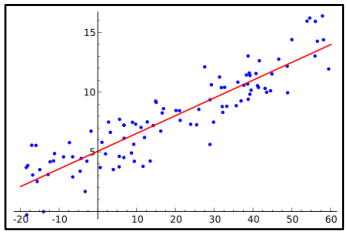

Fig 2. Line of best fit and linear regression (via upgrad)

The line of best fit is based on changing m's and c's values. The discrepancy between the observed and predicted values is known as the predictor error. The values of m and c are chosen so as to provide the least amount of predictor error. It's crucial to remember that an outlier can affect a simple linear regression model. As a result, it shouldn't be applied to large data sets.

The approach of the least squares method is frequently utilized to fit linear regression models, although alternative methods exist. For instance, least absolute deviations regression reduces the "lack of fit" in another norm, whereas ridge regression (which involves an L2-norm penalty) and lasso regression minimize a penalized version of the least squares cost function (which involves an L1-norm penalty). Additionally, the least squares method can be applied to non-linear models. Therefore, even though they are closely related, "least squares" and "linear model" are not interchangeable terms. When the dimensions of the training set exceed the number of data points, the conventional Linear Regression Classification (LRC) approach fails to yield accurate results. [21]

One major application of linear regression is to analyse and comprehend a data set rather than simply predict outcomes. Following a regression, one can learn from the learnt weights how much each characteristic influences the result. For instance, if we have two characterestics namely, A and B that are used to predict a rate of success, and the weight learnt by A is significantly greater than the learned weight for B, this indicates that the occurrence of the trait A is more correlated with success than the occurrence of the trait B, which is interesting in and of itself. Unfortunately, feature scaling undermines this: because we rescaled the training data, the weight for A (a′1) no longer corresponds to A values in the real world.
In general, comparing predictors' (unstandardized) regression coefficients to determine their relative relevance is not a good idea because:

- the regression coefficients for numerical predictors will be determined by the units of measurement of each predictor. it makes no sense, for example, to equate the influence of years of age to centimetres of height, or the effect of 1 mg/dl of blood glucose to 1 mmhg of blood pressure.
- the regression coefficients for categorical predictors will be determined by how the categories were defined. for example, the coefficient of the variable smoking will be determined by how many categories you construct for this variable and how you handle ex-smokers.

1) Comparing standardized regression coefficients

The standardised version of model variables is substituted to derive the standardised regression coefficients.

A variable that has a mean of 0 and a standard deviation of 1 is said to be standardised. Remove the mean and divide by the standard deviation for each value of the variable. The unit of measurement for each predictor changes to its standard deviation when the predictors in a regression model are standardised. We think that by using the same unit to measure each variable in the model, their values will become comparable. Standardized coefficients offer certain advantages, including their ease of application and interpretation, as the variable with the highest standardized coefficient is considered the most significant, and so on. They also provide an objective measure of importance, unlike other methods that rely on domain knowledge to create an arbitrary common unit for judging the importance of predictors. However, standardized coefficients have limitations, as the standard deviation of each variable is estimated from the study sample, making it dependent on the sample distribution, size, population, and study design. Even a small change in any of these factors can significantly affect the value of the standard deviation, leading to unreliable standardized coefficients. In such cases, a variable with a higher standard deviation may have a larger standardized

coefficient, and hence may appear more important in the model, even if it's not.

Comparing the impact of each predictor on the model's accuracy

When using linear regression, you can examine the increase in the model's R2 that comes from each additional predictor or, conversely, the decrease in R2 that results from each predictor being deleted from the model.

In logistic regression, you can evaluate the reduction in deviance that happens as each predictor is included in the model.
The main idea behind comparing predictors is to assess the impact of each predictor in relation to a chosen reference predictor, by comparing the change in one predictor required to replicate the effect of another predictor on the outcome variable Y. This method is particularly useful when a natural reference predictor is available, such as comparing the effects of different chemicals on lung cancer relative to smoking, which can be considered a reference for all lung carcinogens. However, when a natural reference is not available, it is best to use another method to evaluate variable importance.
Another method involves selecting a fixed value or change in the outcome variable Y and comparing the change in each predictor necessary to produce that fixed outcome. For example, becoming a smoker is equivalent to losing 10 years of age in terms of the 10-year risk of death from all causes for a middle-aged man. This method can also be useful in assessing the impact of predictors on the outcome variable Y.
For example, when it comes to the 10-year risk of death from all causes for a middle age man, becoming a smoker is equivalent to losing 10 years of age.

a. Logistic Regression

A statistical model called the logistic model, also referred to as the logit model, estimates the likelihood of an event by converting the event's log-odds into a linear combination of one or more independent variables. Logistic regression, commonly referred to as logit regression, calculates a logistic model's parameters in regression analysis (the coefficients in the linear combination). According to its formal definition, binary logistic regression has a single binary dependent variable (two classes, coded by an indicator variable) with the values "0" and "1," whereas the independent variables can either be continuous variables or binary variables (two classes, coded by an indicator variable) (any real value) [22] [23] .
Only when a selection threshold is included does logistic regression become a classification approach. The classification problem itself determines the threshold value, which is a crucial component of logistic regression. The precision and recall levels have a significant impact on the choice of the threshold value. In an ideal situation, precision and recall should both equal 1, but this is very rarely the case.

In the precision-recall tradeoff, we use these cases:

1) *Low Precision/High Recall:* We choose a decision value that has a low value of Precision or a high value of Recall in applications where we wish to lower the number of false negatives without necessarily reducing the number of false positives. For instance, in a cancer diagnosis application, we don't want any impacted patients to be labelled as unaffected without paying close attention to whether the patient is receiving a false cancer diagnosis. This is due to the fact that additional medical conditions can identify the absence of cancer but cannot detect its presence in a candidate who has once been rejected.

2) *High Precision/Low Recall:* We select a decision value that has a high value of Precision or a low value of Recall in applications where we wish to cut down on false positives without necessarily cutting down on false negatives. For instance, if we are predicting whether a customer will respond favourably or unfavourably to a customised advertisement, we want to be absolutely certain that the customer will respond favourably to the advertisement because a negative response could result in the loss of potential sales from the customer.

The main difference that arise between logistic and linear regression is that the range of logistic regression is contained within a value between 0 and 1. In contrast to linear regression, logistic regression does not require nor does it demand a linear relationship between the input and output variables. This is because the odds ratio were converted to a non-linear log transformation. The mathematical function can be defined as:
Logistic function f(x): $\frac{1}{1+e^{-x}}$

Where x is an input variable

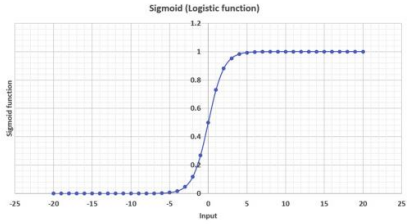

Fig 3: Illustration of a sigmoid function on x-y graph

Even though a number of algorithms, including SVM, K-nearest neighbours, and logistic regression, demand that features be normalised, Principle Component Analysis (PCA) provides a good illustration of why normalisation is crucial. In PCA, the elements that maximise variance are what we are most interested in. If those attributes are not scaled, PCA may find that the direction of maximal variance more closely relates to the 'weight' axis if one component (for example, human height) fluctuates less than another (for example, weight) due to their respective scales (metres vs. kilogrammes). This is obviously false because a change in height of one metre can be thought of

as being far more significant than a change in weight of one kilogramme.

IV. RESULTS

From the ML strategies deployed, we were able to create a Confusion Matrix for the sake of accuracy. Also the value were displayed in percentage for a clearer grasp about the model's reliability in layman's term.

The accuracy varied from 87% to a little more than 94% as evidenced from the result. The 1 in the chart displayed the chances of recession, while a 0 indicated little chances of recession.
Also used were the CFNAI diffusion index for the values till June 2022, after which some discrepancies were noticed from the subsequent releases of the CFNAI datasets.
From Fig 9 and 10, we can see the accuracy of the plotted versus the actual CFNAI Diffusion Index, which (if it differs) has only minute differences between them. The monthly average of 3 plot was similarly plotted using the ML technique using the dataset.

Fig 4. Unscaled Logistic Regression accuracy

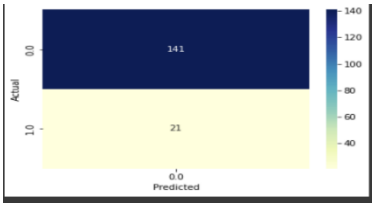

Fig 5. Confusion Matrix for Unscaled logistic Regression

Fig 6. Scaled Logistic Regression accuracy

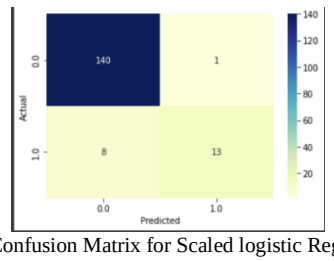

Fig 7. Confusion Matrix for Scaled logistic Regression

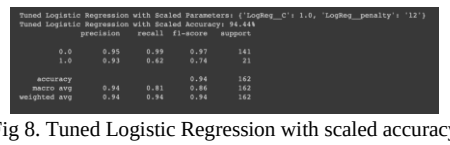

Fig 8. Tuned Logistic Regression with scaled accuracy

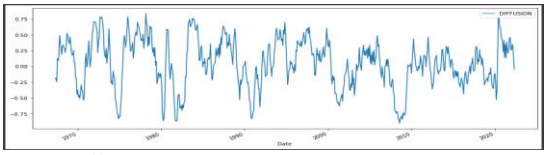

Fig 9. CFNAI Diffusion Index plotted using ML technique. Notice the trough coincides with the period of economic recession.

Fig 10. Actual CFNAI Diffusion Index

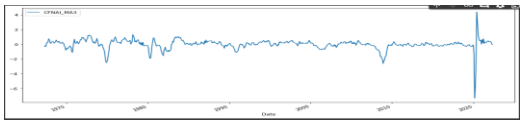
Fig 11. CFNAI MA_3 plotted using ML technique.

From Fig 9 and 10, we can see the accuracy of the plotted graph versus the actual CFNAI Diffusion Index, which (if it differs) has only minute differences between them. The monthly average of 3 plot was similarly plotted using the ML technique using the dataset extracted from the CFNAI. The dataset was taken from the year 1962 to the June of 2022.

V. REFERENCES

[1] M. Brunnermeier (2009) Deciphering the liquidity and credit crunch 2007-08. J. Econ. Perspect

[2] Erkens, D. H., Hung, M., & Matos, P. (2012). Corporate governance in the 2007–2008 financial crisis: Evidence from financial institutions worldwide. *Journal of Corporate Finance*, *18*(2), 389–411. https://doi.org/10.1016/j.jcorpfin.2012.01.005

[3] *Business Cycle Dating*. (n.d.). NBER. https://www.nber.org/research/business-cycle-dating

[4] Chen, J., Vargas-Bustamante, A., Mortensen, K., & Thomas, S. B. (2013b). Using Quantile Regression to Examine Health Care Expenditures during the Great Recession. *Health Services Research*, *49*(2), 705–730. https://doi.org/10.1111/1475-6773.12113

[5] Dabrowski, J. J., Beyers, C., & de Villiers, J. P. (2016). Systemic banking crisis early warning systems using dynamic Bayesian networks. *Expert Systems With Applications*,*62*, 225–242. https://doi.org/10.1016/j.eswa.2016.06.024

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

[7] Nyman, R., & Ormerod, P. (2020). Understanding the Great Recession Using Machine Learning Algorithms. *arXiv*. https://doi.org/10.48550/ARXIV.2001.02115

[8] 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

[9] Paruchuri, H. (2021). Conceptualization of Machine Learning in Economic Forecasting. *Asian Business Review*, *11*(2), 51–58. https://doi.org/10.18034/abr.v11i2.532

[10] Estrella and F. S. Mishkin, "Predicting U.S. recessions: financial variables as leading indicators", *The Review of Economics and Statistics*, vol. 80, no. 1, pp. 45-61, 1998.