Declaration on Plagiarism

Project Title:	Recession Prediction: A Comparative Analysis of Machine Learning Models
Student Id(s):	19210377 and 19211010
Student name(s):	Yangchen Dolkar Sherpa and Garima Singh
Student email(s):	Yangchen.sherpa2@mail.dcu.ie and garima.singh4@mail.dcu.ie
Chosen Major:	MSc in Computing (Data Analytics)
Supervisor:	Dr. Martin Crane
Date of Submission:	17-08-2020

DISCLAIMER

I/We declare that this material, which I/We now submit for assessment, is entirely my/our own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my/our work. I/We understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I/We have read and understood the Assignment Regulations. I/We have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me/us or any other person for assessment on this or any other course of study.

Name(s): Garima Singh & Yangchen Dolkar Sherpa

Date: 16-08-2020

Recession Prediction : A Comparative Analysis of Machine Learning Models

Garima Singh
School of Computing
Dublin City University
Dublin, Ireland
garima.singh4@mail.dcu.ie

Yangchen Dolkar Sherpa
School of Computing
Dublin City University
Dublin, Ireland
yangchen.sherpa2@mail.dcu.ie

***Abstract*—Our research aims to investigate the reliability of different Machine Learning models in predicting US recession. The selected Machine Learning models', namely, SVM, Random Forest, MLP and KNN are evaluated for accuracy at different time periods.**

In this paper, we use economic data-sets from the United States between 1960 to 2020 and check the accuracy of predictions made by the above models one, three and six months prior to the actual occurrence. Various macroeconomic financial indicators were used for our study such as long- and short-term U.S federal government interest rates, Standard and Poor's, unemployment rate, Consumption Index and Federal funds rate. The "recession state", obtained from NBER, was used as the binary dependent variable for the same period. Past studies on recession prediction focused solely on the treasury rates and term spread hence through this study we use other additional relevant macroeconomic indicators and implement them in the various models.

The results show that SVM has the best accuracy for one month ahead prediction while KNN has the best accuracy for six months ahead prediction, with the lowest number of false positives, and can be considered an ideal model for an early warning indicator. Both the models perform accurately for out of sample tests and predict the latest recession that began in 2020 Q1.

***Index Terms*—Random Forest, SVM, KNN, MLP, ROC Cross validation**

I. Introduction

National Bureau of Economic Research [3] describes recession as "a significant decline in economic activity that is spread across the economy and that lasts more than a few months". The recurrent expansion and contraction in the growth of the economy represent a business cycle. The NBER releases the data on peak and troughs during the business cycle where the peak represents the onset of recession [14].

Due to consistent trade wars between countries and the current prevailing pandemic conditions, the global economy is experiencing a slowdown[1], showing how any expected or unexpected risk in the economy is capable of causing a recession. Thus, forecasting recession has become imperative not only for major businesses but even for working professionals to prepare oneself for the coming state of the economy.

[1]https://www.bbc.com/news/business-52972901

This brings in the need for recession prediction models, which when given real-time global indicators can lead to accurate predictions that can help policymakers and the government to come up with solutions before getting hit by it. Additionally, having a model that accurately provides early warning indicators has the potential to give them time to adjust national financial and economic policies.

As the world's largest economy with a high percent of world's GDP accounted from, the United States plays a key role in setting precedent for the international policies and world trade of import-export. Thus, the forecasting abilities of existing machine learning models were investigated based on economic data from the US such as long- and short-term U.S federal government interest rates, Standard and Poor's (S&P), Federal funds, unemployment rate, consumption indices and FED funds rate. The data-set was obtained from FRED and was sampled monthly. As the predictive power of the yield curve has been recognized in many previous studies [6][7], it became a high priority feature. The short-term bonds give various maturities' returns and long-term bonds provide expectations for future economic growth. Due to multiple studies claiming to include different groups of features such as output and incomes, Labor market and interest exchange rates, the report further discusses the merit of individual features with the upcoming recession prediction and forecasting. Including the variables with their individual lags of three, six and twelve months leads to more efficient forecasting of the output.

There have been influential reviews from the past literature about the forecasting accuracy, with few quarters ahead predictions by using the yield curve to predict GDP. The Gross Domestic Product was initially taken into account in this research to study the pattern of recession and proofreading the prediction result afterwards. As it has been observed that GDP began to fall in the third quarter of 2008 and was not resumed until the third quarter of 2009, which is in line with the actual recession and hence we considered the data from NBER to capture the recession state after matching it with the pattern in GDP.

The methodology in the report can be distinguished by the use of linear, non-linear models and neural network models. Support vector machine (SVM) is linear in nature as compared to Random Forest Classifier (RF) which is non-linear, Multilayer Perceptron (MLP) and K-Nearest Neighbour (KNN) as neural networks. Non-linear models are acceptable replacements to linear regression models due to being unbiased and flexible[2] but they do have drawbacks. Researchers have included SVM and RF multiple times with linear, probit regressions and MLP in past works, thus an addition of KNN model was analysed in the paper to get a better perspective through neural analogy.

The purpose of the paper is to examine the accuracy level of different models on the quantitative and categorical data-set. The prediction of US recession one month, three months and six months ahead are performed to examine all models' individual prediction quality. The primary challenge was the selection of macroeconomic variables and their predictability on the US recession. The data obtained from FRED was pre-processed further. K fold cross validation was used in the training data ranging from 1960 to 2008 and the data from 2009-2020 was used for validation.

This research aims to answer the following questions; How well did the models perform in forecasting US recession one, three or six months ahead of the actual event? Which model performs precisely in different economic prospects and Which model is most prone to generating false alarms before US recession?

This report is structured as follows: Section 2 is literature review for the past work done on recession prediction with different data-sets and model applications, Section 3 conveys the data-set, methodology used on the pre-processing of data, Section 4 is the result and analysis observed from the research performed on the comparative analysis of different models for the data-set going back to 1960s, the last section is the final conclusion with references.

II. Literature Review

Estrella and Mishkin [2] in their paper of Predicting US recessions: Financial variables as leading Indicators, played an instrumental role in selecting the financial indicators. The work concludes that S&P 500 index provides additional information for recession prediction which is not present in yield curve spread. Additionally, other macroeconomic indicators, selected for the study, are good for one quarter ahead forecasting. It states that macroeconomic indicators are most prone to over-fitting. Hence the indicators were selected accordingly. We see that in most past studies, the yield curve is the primary focus for predicting recessions. Additionally, the extensive study by Stock and Watson [3] provides a thorough idea about the various leading economic indicators and how they affect future economic growth and its correlation with recession. According to Nyberg, H. [4] in their article on Dynamic Probit models and financial variables in recession forecasting in 2010, stock market returns have additional predictive power in addition to yield curve as was stated previously by Estrella and Mishkin. The method of adding lags on the dependent variable was used where the lag is equal to the forecast horizon. They state that with longer lags the predictive power seems to diminish.

S. Kozicki [6] discusses the effect of yield spread [3] and yield curve on the prediction of inflation and real growth in his paper of Predicting real growth and inflation with yield spread. The result of yield spread has maximum predictive power for inflation and can predict three years beforehand whereas for real growth it can predict for the next year or so. It also states that the level of short rates matters for predicting inflation whereas the yield spread matters for predicting real growth.

In 2006, Wright [17] in his article uses the Probit model for forecasting recession and states that the usage of Federal Funds rate and term spread gives better predictive performance than using the term spread alone. The term spread helps to explain the shape of the yield curve. The low term premium represents the flatness of the yield curve hence using only yield curve as an indicator may not produce accurate results. This is further researched in the following paper.

Since most papers used Treasury Bill rates as the indicator for the research one might think that it is sufficient. However, the claim is refuted by Fintzen and OStekler [1] in their research on Why did forecasters fail to predict the 1990 recession. From their study, a deduction was made that most models failed to forecast the recession in 1990 due to the limited number of economic variables that were used. Understanding the economic variables and selecting them while avoiding over-fitting is necessary for recession prediction. Klein and Moore [11] explain the different indicators used in economic forecasting and sheds light on the performance of the different background and group indicators in practice. Their research played a major role in selecting the financial variables used in our study.

In terms of the machine learning models to be trained in the study, there has been a close relationship established in past literature on which model is familiar in recession forecasting. Gogas, et al. [12] wrote a research paper of Yield Curve and Recession Forecasting in a Machine Learning Framework, where recession prediction was done using SVM classifier by using GDP as the dependent variable and long and short term US federal government interest rates as the independent variable. It uses k-fold cross validation to avoid over-fitting. Overall, it produces 74.2% overall accuracy and 6 cases of

[2]As conceptualized by Jim Frost in the blog https://statisticsbyjim.com/regression/choose-linear-nonlinear-regression/

[3]As stated in Investopedia article, the yield spread is difference between yields on the differing debt of varying maturities, risk level or credit associated.

false positive results. Similarly, Nyman and Ormerod[13] used long and short-term yields and S&P 500 as independent variables and GDP as the dependent variable. They used Random Forest and Ordinary Least Square Regression models to assess its potential for early warnings of recession. The results gathered from their study show that Random Forest accurately predicted recessions 3 quarters ahead and is capable of predicting 6 quarters ahead but slightly later than the actual event. Thus, confirming SVM classifier and Random Forest wide usage in various studies and research with their predictive capabilities for economic training is admirable.

Furthermore, Nyman and Ormerod [15] in the study of the Great Recession attempt to predict the 2001 recession by replicating the actual forecasting situation. They use 3 year and 10-year US government bonds, S&P 500. The results obtained from the Survey of Professional Forecasters (SPF) are compared to the results from the Random Forest model and one step ahead forecast is better through SPF while the performance of Random Forest is better when predicting four quarters ahead. It sheds more light on Random Forest's long-term prediction power and timeliness. Like Random Forest and SVM, different neural networks and linear regression models are applied in research of Giannopoulos and Aggelopoulos [10] for Predicting SME loan delinquencies during recession using accounting data and SME characteristics in 2019. They concluded that in terms of short and accurate prediction MLP neural network works better than others and SVM have better effectiveness regarding average accuracy for long term predictions.

Filardo [5] on the reliability of recession prediction models, associates the consistency of five models are researched namely: simple rules of thumb using the Conference Board's composite index of leading indicators (CLI), Neftçi's model, Probit model, GDP forecasting model, and Stock-Watson predicted model, their prediction is based on the accuracy and timeliness of the models. Even though some false signals were generated by each model it produces reliable results. This theory is compared was further compared in his [14] paper discussing US 2001 recession. The paper's result with its early warning indicator being accurate for both models is compared and issue with real-time sensitivity is a concern, further tells us that indications received are better if all models are in agreement. The latter research mainly describes the non-parametric CLI rules of thumb and Neftci provides 4-8 months warning prior recession.

In the paper of What does the yield curve tell us about GDP growth? by Ang, et al. [7] in 2006, imposes the absence of arbitrage in bond markets in order to develop a yield model to predict GDP growth by taking few yield and GDP growth as state variables. It also states the advantages of using yield models over unrestricted Ordinary least squares (OLS) regression specifications. It specifies that the model is useful in choosing the right spread maturity to predict GDP growth such that the best slope is obtained from maximal maturity difference. The slope of the yield curve in forecasting GDP growth is affected by the nominal short rate. It states that the implemented term structure model is comparable to the Stock-Watson indices model. For short term predictions both generate good results while for long term prediction, the term structure model performs better.

Additional on yield curve's benefit on the recession, the research on the forecasting yield curve using Artificial neural networks by Nunes .M, et al [16] sheds new light on the behaviour of the ANN models on economic indicators. The models used for the study were Multivariate linear regression and Multilayer perceptron model (MLP). It states that since the order in which the data is placed is relevant it generates new features from the current features using lagged values in order to incorporate the past values of the time series data into the models. It concludes that MLP produces a higher level of accuracy in predicting the yield curve. Data pre-processing methods mentioned in their research have been proven valuable.

While the prediction accuracy of linear models is better when "recession state" is used as a binary dependent variable rather than using output growth of GDP as the dependent variable when the independent variable is the yield curve. This hypothesis is stated in the report by Chauvet and Potter [8] of predicting recession based on the yield curve's structural breaks. The effect of structural breaks on the probability of recession is tested using the Probit model. It concludes that each business cycle is different hence structural breaks play a major role while predicting recessions.

These research papers discuss the various recession prediction models namely: SVM methodology, Probit model, Logit model, Random Forest, Yield model, MLP and Gradient Boosting methods for various scenarios. Thus, it helps to understand which algorithm to use and what features to apply them to.

The research will compare the accuracy of Random Forest Machine Learning, SVM, MLP in addition to KNN. From the features carefully chosen in the past research, our study has incorporated additional macroeconomic features, thus we also use the "Recession state" as the binary dependent variable for better prediction (Chauvet and Potter) and add subsequent lags to the current indicators as new features (Nunes, Gerding, McGroarty and Niranjan).

III. Methodology

A. Data Specification

Taking into account the impact of the chosen variables on economic growth, the indicators that were chosen were long term and short term U.S federal government interest rates,

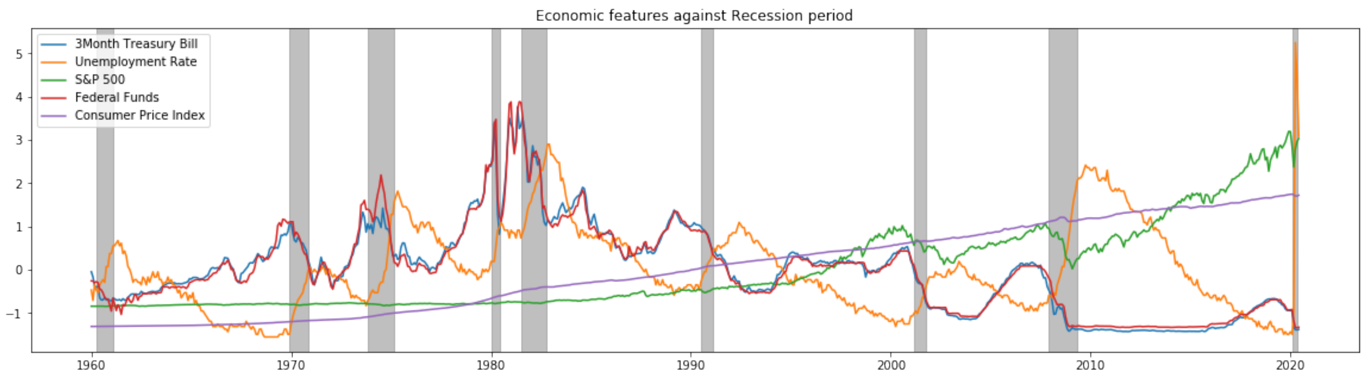

Fig. 1. Economic features against Recession Period

S&P, Federal Funds, unemployment rate, Consumption Index and Fed Funds rate. The binary dependent variable chosen for our study was 'Recession State' which varies for each of the above independent variables used that are used as indicators.

The data spans from 1960 to 2020, sampled on a monthly basis, and were extracted from the database of the Federal Reserve Bank of Saint Louis, Federal Reserve Economic Data (FRED) [4] and Yahoo Finance[5]. The Recession state in the given dates was obtained from the National Bureau of Economic Research (NBER).

Both short term and long term Treasury yield data were used for this study. The short term yield used had a maturity of 3 months and 6 months and the long term yield used had a maturity of 1 year, 5 year and 10 years, where maturity is the end of transaction bond or interest rate life. Treasury yield is an investment or an interest rate a government pays. Each yield has a different maturity rate ranging from 1-month to 30-year and the yield curve is the relationship between yields and maturities. The yield curve changes daily as the investors review economic situations and future bonds [6]. If the slope becomes negative as the curve inverses, such as short-term borrowing costs becoming higher than long term loan costs, gives away an expectation of weak economic growth. This curve helps monitor recession as it is known to be highly sensitive.

S&P 500 indicates the market value index of 500 largest traded companies based on their size, liquidity and index, giving the accurate stock market representation to the public. The steep declines in the stock market and recession making them parallelly related [7]. The Federal changes interest rates to adjust the supply of credit in the economy, i.e. monetary policy. They usually try to bail out the borrowers, such as banks, which in turn lead to low interest rates on loans and bonds to consumers. This affects businesses and consumers getting lower interest rates on savings. Consecutive quarters negative economic growth leads to recession, one of the factors measuring it is the unemployment rate. These two are usually collateral [8], the main focus being businesses lagging hiring, adding to unemployment.

The Federal funds rate[9] is described as the interest that is charged across banks for lending the money from the respective reserve balances on an overnight basis. The Federal funds rate is considered one of the most important interest rates in addition to the yield curve and has a huge effect on monetary and financial conditions. The FOMC (Federal Open Market Committee) is responsible for implementing the monitory policy. Hence, it goes highly correlated with the recession.

In addition to the above mentioned indicators, Unemployment rate and consumption index were selected. A Recession occurs when there is negative economic growth for two consecutive quarters which is often shown by various indicators one of them being the rise in the unemployment rate and decline in consumption index. Thus, the two features provide an idea of the impending recession.

Thus, prediction models use a handful of features to create more stimulating results. One of the main challenges that are faced prior to prediction is selecting the most suitable features based on the models that are to be used. To avoid model failure and over-fitting, one has to be specific in updating the features' significant lags and values in pre-processing, to make them best suited in fitting each model.

[4]https://fred.stlouisfed.org

[5]https://finance.yahoo.com/

[6]referred as CMTs or Constant Maturity Treasure, the index based on the average yield of different maturities of different stages

[7]Royal Bank of Canada shows research related to decline of S&P in recession market based on history, with plunges of 24% and 32%

[8]Based on a study of Investopedia, during recession companies faces falling revenue thus unemployment rises

[9]As per Investopedia article in monetary policy https://www.investopedia.com/articles/investing/081415/understanding-how-federal-reserve-creates-money.asp

B. Data Exploration and Pre-Processing

In this section, the indicators that are used for our study is further explored to justify its effects on recession and is plotted in Fig 1. The graph is designed through standardized data to show the relationship of various indicators to recession. The following findings were noted during data exploration:

- During the 1960's recession, there was a sharp increase in unemployment and a decline in treasury bills, CPI, stock price index and fed funds rate.
- Similarly, for 1970 the two recessions were caused by the increase in oil prices.
- There is a slight downward curve before the recession in treasury bills, CPI and fed funds rate and a slight upward curve in unemployment.
- In the 1980s There was a sharp increase in treasury bill rate due to high inflation and heightened nuclear fear. The Treasury bill rates were declined and stabilized to more normal rates in the late 1980s due to lower inflation and decline in marginal tax rates.
- The recession in 1990 was caused by a sharp fall in consumption after the uncertainty raised by Iraq's invasion of Kuwait which led to a sharp increase in oil prices. According to the literature, several forecasters failed to predict the 1990's recession since they relied on a small number of indicators.
- Rightly so, the graph shows an upward turn in treasury bills, Fed funds rate before the recession which is highly misleading hence adding more features to the model will help with better accuracy.
- The great recession of 2007 to 2009 perhaps provides the clearest view of the data with the unemployment rate rising slightly and the slow decline in the treasury bill rates, Fed Funds rate and Consumer Price Index before the recession effectively present the relationship between the selected indicators and impending recession.

The latest recession of 2020 starting from February caused by the global pandemic COVID-19 [10] shows a sudden increase in unemployment but the other indicators are stable. Hence predicting the recession of 2020 12 months or even 6 months prior is not possible because of the unexpected nature of the event. A stable increase in S&P 500 and CPI can be observed over the years with minimal fluctuations hence these are selected as well as stable parameters for our model.

Lags of one, three and six months were added to the leading indicators of recession in order to obtain the result in different timestamps and add additional features for the models hence there were a total of 65 independent variables. Recession state, obtained from NBER, was utilized as a binary dependent variable which was classified into two classes: 1 indicated the occurrence of recession and 0 indicated that there was no recession in that month. A shift of 1 month, 3 months and 6 months were done on recession in order to understand the early warning signs through dependency as well.

The economic data is released at different frequencies hence the data considered were sampled monthly [11]. In order to make the data consistent with the same timestamp, the historical data of a few columns had to be removed at the beginning of the period and the end since additional rows was added as lags and shifts were performed on the data-set. The data were standardized using skit-learn library in python [12] to transform the data into a uniform scale since all the indicators had different units. The computation is done using this library calculated using z-score internally. The models that were implemented therefore dealt with consistent data.

Furthermore, 80% of the clean data was separated into training set i.e. the data from 01-01-1960 to 04-01-2008, where the models were trained for estimation, and 20% into testing set i.e. 05-01-2008 to 05-01-2020, where the performance of the model was evaluated based on the training set.

C. Models

The models mentioned in the literature review are further discussed in this section. While the first three models were selected based on past research, KNN is added due to having outperforming results from major classifiers when used in economic forecasting.

SVM [10], which is a supervised machine learning technique-a binary classifier, is most useful for predictive analysis on labelled categories with two class data classification as we have 'Recession state' which is the binary dependent variable. SVM uses a hyper-plane to separate the data points[13], known as Support Vectors shown in Fig 2, into positive and negative classes. The distance between support vectors is maximized to find suitable hyper-plane in order to avoid overfitting.

Similarly, Random Forest Classifier is highly popular for binary classification. It is an ensemble-based learning classifier that is modelled by selecting random samples from the dataset which is then segregated to random subsets of decision trees to avoid overfitting. Each decision tree produces a result, each predicted result is then considered for a vote and the prediction result with the greatest number of votes is considered as the final result. The result produced is highly accurate and robust since multiple trees are involved in this process. Since it considers the average of all the results it does not suffer from overfitting. In Random Forest Classification, due to the

[10] As stated in Financial Times and International Monetary Fund, recession occurring due to economic consequences of pandemic

[11] The individual data needs to standardized and combined together in order to implement further, thus due to limited data-set for few features, they were monthly incorporated

[12] A pre-processing package in skit-learn library of python, standardization required when the individual features are not normally distributed

[13] As per Scikit-learn Machine Learning in python, a hyper-plane is an infinite dimensional plane used for classification and regression by achieving separation for different class data points

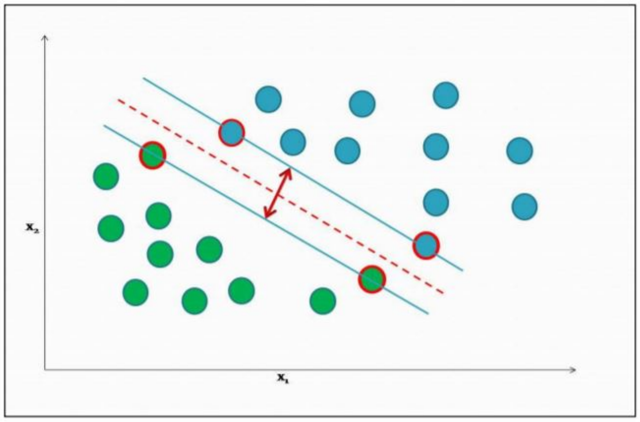

Fig. 2. Hyper plane selection and support vectors. The SV's are represented with the pronounced black contour, the margin lines are represented with the continuous lines and the hyper plane is represented with the dotted line [12]

number of decision trees' contribution, makes it robust and high in accuracy. Scikit-Learn adds variables to the model, which shows the importance of each independent variable used for prediction.

The features used were found to be largely uncorrelated with each other since they belong to different economic research categories[14], thus individual decision trees were created to predict distinct classes of the outcome. Due to the different timeline and groups, the features were unrelated, thus inconsistent. Random Forest classifier deal with such kinds of data via bagging. Ensemble method then produces most accurate predictions from individual model trees created, having low correlation among themselves.

KNN, a non-parametric approach for classification or regression, is highly effective for classification problems with low dimensional such as economic data [18]. The nearest neighbours from the training data set are selected in order to classify each data point. The value of the predicted point is obtained by averaging the distance between the K points which were weighted equally. The distance metric that defines the distance between the data points is highly sensitive hence it does not work well for problems with many irrelevant predictors. Hence, the limited indicators have been selected according to the literature.

In the Multilayer Perceptron neural network [10], one of the main reasons is to work on classification prediction problems. In time series analysis, MLP, when used as a baseline for the recession prediction, is best suited for further comparison with different models.

As MLP consists of three or more layers: input layer, one or more hidden layers and output layer. The basis of architecture used in the report was 2 hidden layers with 5 and 2 hidden units for the diversified independent features used, giving away the number and size of hidden layers as shown in Fig 3. We train the neural network by lbfgs, a quasi-Newton optimizer for solving the classification prediction, which is used due to several independent data points in the study and lbfgs convergence with limited data-set. Equivalent numbers of false positive and negative were discovered, giving a baseline model score for others. Individual cross-validation accuracy scores were obtained to analyse the performance of the models in different time spans. Usually, classification accuracy is not enough to decide whether the model is good enough, thus precision and recall performance measure is also taken into account. This was further understood and analysed from a confusion matrix, detailing the best accuracy in a model with a narrative behind it.

[14]https://fred.stlouisfed.org/categories

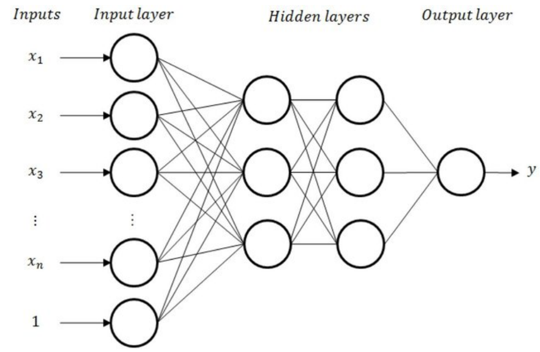

Fig. 3. Multilayer Perceptron with two hidden layers by Guérillot [19]

Comparison of the results was obtained by cross-val ROC scores and F1 scores for different models. The method is performed in four steps: training, fitting, predicting and testing. The largest part of the data-set is used as training data from 1960 to 2008 and the rest from 2008 to 2020 is used for out-of-sample testing.

ROC was used as a performance metric for the models using K fold cross validation. Understanding the true performance of the model is essential since a high score generated from training and testing will be biased as it would be unknown whether the score is a result of over-fitting or true performance hence we use cross validation technique to avoid over-fitting and observe the performance of the model using the validation set. The data-set is split randomly to k folds and the training and testing set is repeated k times. A different set of k is used as a test set while the rest of k-1 is used as a training set. The final evaluation of the model is obtained by averaging the performance of the model on every fold. The value of k was set to 3 in this scenario.

A ROC curve shows the relationship between the False Positive rate and True Positive Rate, which are the specificity and sensitivity of the model created. The equations used for this curve are shown in Equation 1 and 2.

$$True\,Positive\,Rate = \frac{True\,Positive}{True\,Positive + False\,Negative} \quad (1)$$

$$False\,Positive\,Rate = \frac{False\,Positive}{False\,Positive + True\,Negative} \quad (2)$$

The imbalanced class distribution among the features used in the process requires F1 score, which is beneficial in real-life scenarios. Precision is the number of positive predictions over total positive class values, called the positive predictive value. While the recall is positive predictions over the total number of positive class values in the test data-set. F1 is the combination of precision and recall, such that the false positive and negative can be balanced and conveyed properly from a model prediction. These positive and negative rates are used for calculation of precision and recall in F1 score and further in evaluation, shown in Equation 3. From the results, there were many false measurements noted, thus F1 was predominant in the analysis.

After evaluating the two scores the best 2 models that performed best in the indicated time were selected as a baseline model and then the probability of the occurrence of recession is predicted using the validation set.

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

IV. Results and Analysis

In this section, the results obtained from the four models were discussed in terms of cross validation ROC score accuracy, F1 score accuracy, the true positive and true negative rate of the models and the comparison of models based on early warning signals. The figures Fig 4, Fig 5 and Fig 6 were plotted against the actual values and the predicted values from out of sample validation, where the grey area denotes the period of recession according to NBER. The in-sample cross validation scores obtained from the training data is depicted in Tables I, II and III. The confusion matrix was prepared in order to find the true positive, true negative, false positive and false negative rates from the predicted results.

Models	ROC Score	F1 Score
One Month Ahead		
RF	0.947	0.839
SVM	0.937	0.899
KNN	0.811	0.909
MLP	0.848	0.865

TABLE I
Measurement Scores for 1 month ahead prediction

Models	ROC Score	F1 Score
Three Month Ahead		
RF	0.899	0.844
SVM	0.845	0.825
KNN	0.737	0.887
MLP	0.761	0.875

TABLE II
Measurement Scores for 3 month ahead prediction

Models	ROC Score	F1 Score
Six Month Ahead		
RF	0.889	0.847
SVM	0.757	0.835
KNN	0.682	0.858
MLP	0.666	0.852

TABLE III
Measurement Scores for 6 month ahead prediction

The cross validation F1 and ROC scores of the models were then compared for our in-sample analysis for the data ranging from 1960 to 2008 to find the two best performing models to be used for the prediction. The models selected from the in-sample analysis were used to perform out of sample testing on the validation data ranging from 2009 to 2020.

Considering the indicator of recession 6 months in advance, Random Forest generates a ROC accuracy of 0.889 and KNN leads the F1 score by 0.858. These two models were selected as the baseline models for the prediction of the earliest warning indicator, as shown in Fig 6 both the models performed fairly well from the 1960's to 1975 and captured the recessions. However, in 1980 Random Forest performed slightly better. Subsequently, in 2009 both the models generated false positives, while the performance of KNN stabilized from 2010 Random Forest generated false positives since. Hence for the earliest warning indicators, Random Forest generated 77 false positives and 6 false negatives while KNN generated 12 false negatives. KNN produced 128 true positives and Random Forest produced 51. Hence, we can conclude that KNN has a high true positive rate and performs better in this scenario.

The models selected as baseline models for 1-month prior prediction were SVM, Random Forest and KNN. While SVM and Random Forest had ROC accuracy of 0.937 and 0.947 each, KNN had an F1 score accuracy of 0.909. The Fig 4 represents the prediction probability of the three models. The three models accurately predicted recession up till the 1980s after which KNN showed a negligible false positive peak. The prediction was stable until 2010 when Random Forest started showing false positive from 2015 to 2019. The sharp increase in 2020, the recession caused by COVID-19, is clearly identified by the three models. The performance of SVM and KNN are comparable while Random Forest showed 5 false positives. We can conclude that SVM performed best in short

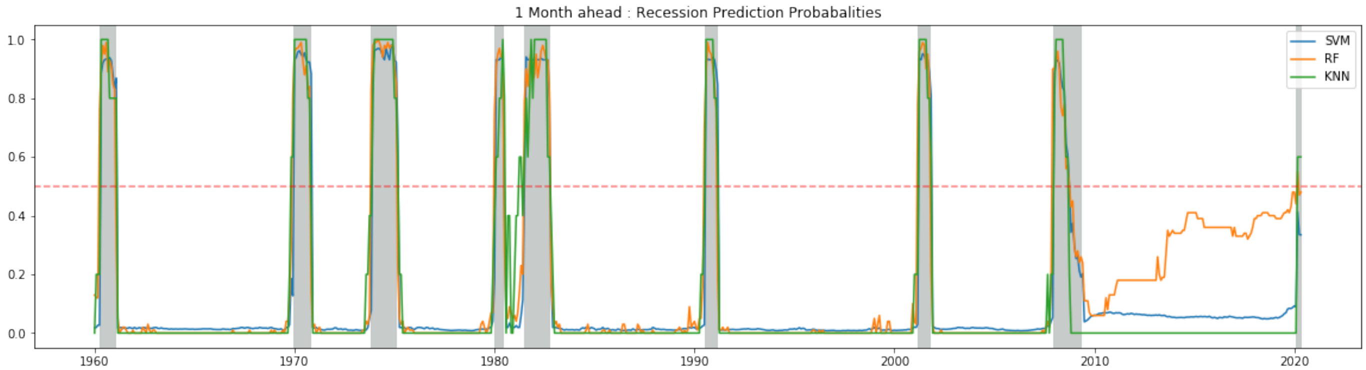

Fig. 4. Recession Prediction probabilities of models : 1 Month ahead

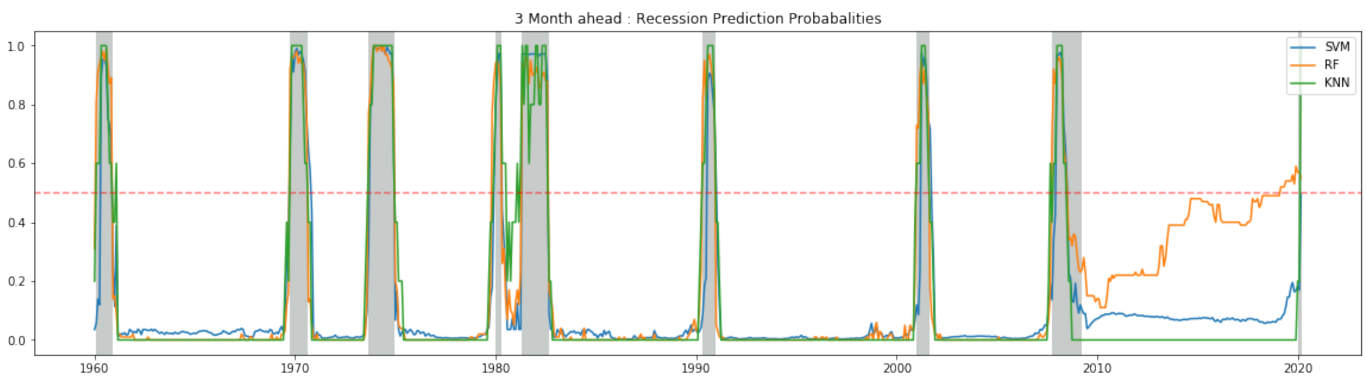

Fig. 5. Recession Prediction probabilities of models : 3 Month ahead

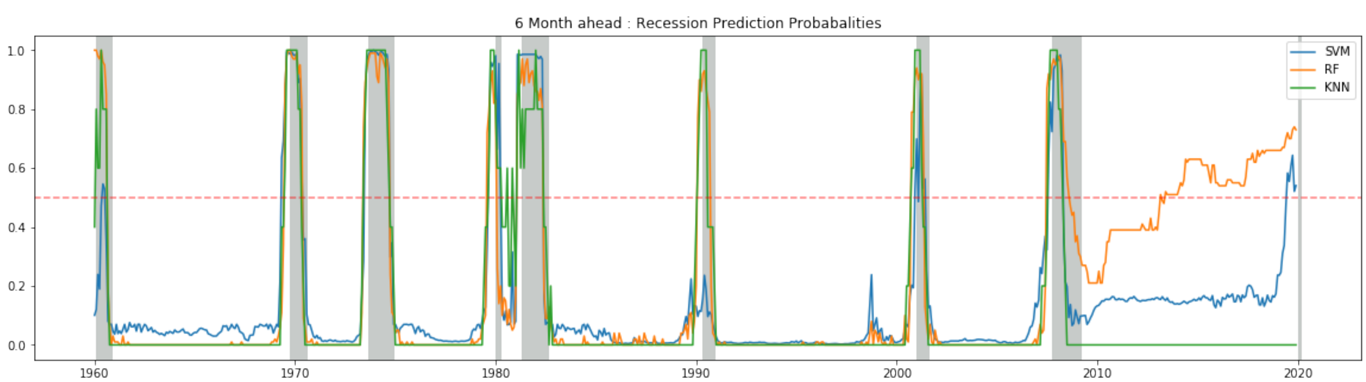

Fig. 6. Recession Prediction probabilities of models : 6 Month ahead

term indications. Even though KNN had the same number of true positives (128), SVM clearly identified 10 true negatives while KNN identified 9.

The research also explores which model has better predictive power 3 months in advance. In this event, Random Forest and KNN had better accuracy at, 0.899 ROC score for Random Forest and 0.887 F1 accuracy for KNN. According to Fig 5, the results were somewhat similar to Fig 6 where KNN produced some irregularities and false positives in the 1980's while the performance of all the models was stable till 2009 after which Random Forest produced false positives till 2019. Random Forest produced 42 false positives while SVM and KNN produced none. The performance of both SVM and KNN is comparable and it can be concluded that both the models produce good results 3 months ahead of a recession.

It is also significant to note that while Random Forest produces comparable early warning indicators, it also produces the greatest number of false positives. Hence, during the validation stage, the accuracy of the Random Forest is significantly low.

Finally, KNN Produces better early warning signs while Random Forest follows close behind. SVM can predict the event most effectively and hence is a good model if one needs to select the model that provides the latest warning indication. Although MLP was not selected for out of sample prediction it is a good baseline for the comparison of models. The context of the recession has been different each time in history and the economy is ever-changing hence it is difficult to forecast the exact time period of recession. As identified in the latest recession of 2020 for example, one can see that all the models effectively predict the recession if it is one month ahead whereas it is difficult to predict the same if the early warning indicator of 6 months ahead is considered.

CONCLUSION

Based on our study KNN produced the best accuracy for early warning signals which contradicts Nyman and Ormerod who state that Random Forest can produce early warning indicators for predicting recessions [11]. Furthermore, when we consider the output for the latest warning signal, all the models perform well however, SVM is the best performing with an in-sample accuracy of 94%. It effectively captures all the recessions and does not generate any false positives. For 3 months ahead prediction we see that the performance of KNN and SVM are similar. The results based on in-sample analysis show that the performance of the models in all horizons was better than MLP hence it was not selected for out of sample prediction.

Even though Random Forest is a good predictor for early warning signals which follows close behind KNN, it produces the greatest number of false positives hence the model's reliability in producing accurate results without false signals deserves further research.

In resulting measurements, the high performance of SVM and K-Nearest Neighbour models provides good long term and short-term prediction results. Considering the fact that recessions are rare events and the economic data keeps evolving and are prone to sudden fluctuations one must bear in mind that the prediction of future recessions will not be similar to past recessions hence selecting the relevant macroeconomic indicators for that specific period is imperative. It would be interesting to analyse the performance of the models in different regions of the world and compare the results as to how the models perform from region to region given the selected parameters. Furthermore, an ensemble of KNN and SVM would provide substantive results for both early and late warning systems as observed from their measurements.

REFERENCES

[1] Fintzen, D. and Stekler, H.O., 1999. Why did forecasters fail to predict the 1990 recession?. International Journal of Forecasting, 15(3), pp.309-323. SJR: Q1.

[2] Estrella, A. and Mishkin, F.S., 1998. Predicting US recessions: Financial variables as leading indicators. Review of Economics and Statistics, 80(1), pp.45-61. SJR: Q1.

[3] Stock, J.H. and Watson, M.W., 1988. A probability model of the coincident economic indicators (No. w2772). National Bureau of Economic Research. SJR: Q1.

[4] Nyberg, H., 2010. Dynamic probit models and financial variables in recession forecasting. Journal of Forecasting, 29(1-2), pp.215-230. SJR: Q1.

[5] Filardo, A.J., 1999. How reliable are recession prediction models?. Economic Review-Federal Reserve Bank of Kansas City, 84, pp.35-56. SJR: Q1.

[6] Kozicki, S., 1997. Predicting real growth and inflation with the yield spread. Economic Review-Federal Reserve Bank of Kansas City, 82, pp.39-58. SJR: Q1. SJR: Q1.

[7] Ang, A., Piazzesi, M. and Wei, M., 2006. What does the yield curve tell us about GDP growth? Journal of econometrics, 131(1-2), pp.359-403. SJR: Q1.

[8] Chauvet, M. and Potter, S., 2002. Predicting a recession: evidence from the yield curve in the presence of structural breaks. Economics Letters, 77(2), pp.245-253. SJR: Q1.

[9] Biau, G., Devroye, L. and Lugosi, G., 2008. Consistency of random forests and other averaging classifiers. Journal of Machine Learning Research, 9(Sep), pp.2015-2033. SJR: Q1.

[10] Giannopoulos, V. and Aggelopoulos, E., 2019. Predicting SME loan delinquencies during recession using accounting data and SME characteristics: The case of Greece. Intelligent Systems in Accounting, Finance and Management, 26(2), pp.71-82. SJR: Q1.

[11] Klein, P.A. and Moore, G.H., 1983. The leading indicator approach to economic forecasting—retrospect and prospect. Journal of forecasting, 2(2), pp.119-135. SJR: Q2.

[12] Gogas, P., Papadimitriou, T., Matthaiou, M. and Chrysanthidou, E., 2015. Yield curve and recession forecasting in a machine learning framework. Computational Economics, 45(4), pp.635-645. SJR: Q2.

[13] Nyman, R. and Ormerod, P., 2017. Predicting economic recessions using machine learning algorithms. arXiv preprint arXiv:1701.01428.

[14] Filardo, A.J., 2004. The 2001 US recession: what did recession prediction models tell us?. Bank for International Settlements (BIS) - Monetary and Economic Department, Paper no.148.

[15] Nyman, R. and Ormerod, P., 2020. Understanding the Great Recession Using Machine Learning Algorithms. arXiv preprint arXiv:2001.02115.

[16] Nunes, M., Gerding, E., McGroarty, F. and Niranjan, M., 2018. Artificial Neural Networks in Fixed Income Markets for Yield Curve Forecasting. Available at SSRN 3144622.
[17] Wright, J.H., 2006. The yield curve and predicting recessions.
[18] Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3), pp.175-185.
[19] Guérillot, D.R. and Bruyelle, J., 2017, March. Uncertainty assessment in production forecast with an optimal artificial neural network. In SPE Middle East oil & gas show and conference. Society of Petroleum Engineers