ISSN: 2769-545X

Review Article

Advances in Machine Learning & Artificial Intelligence

Recession Prediction Using Multiple Machine Learning Methods and Historical Economic Data

Philip Mackay[1*] and Mubeen Ghafoor[2]

[1]*School of Computer Science, College of Science, University of Lincoln, United Kingdom*

[2]*Senior Lecturer in Data Analytics, De Montfort University, United Kingdom*

[*]Corresponding Author
Philip Mackay, School of Computer Science, College of Science, University of Lincoln, United Kingdom.

Submitted: 2023 Dec 01; **Accepted**: 2023 Dec 22; **Published**: 2023 Dec 26

Citation: Mackay, P., Ghafoor, M. (2023). Recession Prediction Using Multiple Machine Learning Methods and Historical Economic Data. *Adv Mach Lear Art Inte, 4*(2), 110-122.

Abstract

This study explores the application of machine learning methods to enhance economic recession prediction in the UK and USA, considering the limitations of traditional methods. Various models, including Logistic Regression, Linear Discriminant Analysis, K Nearest Neighbors, Decision Tree Classifier, Gaussian Naive Bayes, Support Vector Classifier, Neural Network, RTC, Long Short-Term Memory, Convolutional Neural Network, and XGBoost, were assessed using economic data since 1900. The UK data encompassed GDP, unemployment rate, inflation, FTSE 100 index, yield curve, and debt levels, while the USA utilized the 50-day simple moving average of 10-year treasury rates minus the 50-day simple moving average of 3-month treasury rates. Performance evaluation involved averaged F1, recall, and accuracy over 100 iterations, with confusion matrices illustrating model predictions against actual events. Long Short-Term Memory excelled with recall and F1 values of 0.96 and 0.97, accurately identifying 11 in 12 Positive USA events. K Nearest Neighbours, Decision Tree Classifier, Random Forest Classifier, and XGBoost demonstrated good results, with recall ranging from 0.99 to 0.75, F1 from 1.0 to 0.69, and correctly identifying 2 in 3 Positive events. Conversely, Logistic Regression, Gaussian Naive Bayes, and Neural Network exhibited less reliable predictions. Linear Discriminant Analysis, Support Vector Classifier, and Convolutional Neural Network were completely inadequate. Using recent data, most models predicted the USA avoiding recession in 2023-24, but the probability increased to 0.5 by mid-2023, then decreased. Logistical Regression, Linear Discriminant Analysis, and Long Short-Term Memory initially predicted no recession, but the probability rapidly increased to between 0.83 and 0.97 by April 2024. While recession avoidance is plausible, modelling indicates an escalating risk. The results underscore the utility of machine learning in recession prediction, emphasizing the importance of diverse training datasets. Algorithmic performance varied, with neural network models, particularly Long Short-Term Memory and XGBoost, proving most accurate. Further enhancements in performance necessitate refining training datasets and leveraging advanced models like Long Short-Term Memory.

Keywords: Machine Learning, Economic Recession Prediction, GDP, Accuracy, Multiple Models

1. Introduction

Over the last two decades, the UK has seen a total of 3 recessions, resulting from differing causes [1]. The 2008 housing crisis, caused by the sudden correction of an unsustainably rapid rise in house prices and unusually low interest rates led to multiple banks collapsing with governments having to intervene to maintain the monetary systems [2]. During the 2020 Covid pandemic, lockdown halted global economies for many months. Despite Government furlough schemes, there was a surge in company failures and redundancies. Government debt soared and consumer demand fell, causing a vicious cycle of further reductions in production, redundancies and reduced incomes, leading to recession [3].

By 2022 economic growth remained low, with the global economy struggling to recover from the long-term effects of the previous recession. The UK government's announcement of a radical change in fiscal strategy with the Sept mini-budget caused a loss of market confidence [4]. The value of the pound and stock markets plummeted. The Bank of England had to intervene, buying government bonds to stabilise the situation and the IMF reported that the UK economy had stalled. In early 2023 the Russian invasion of Ukraine resulted in mass sanctions from European and

North American countries, causing gas prices to soar in Europe and wheat-based food prices to increase throughout the world. This sharp rise in fuel and food costs caused inflation and wage demands to soar and consumer demand to dip, increasing the risk of a major recession [5].

The United Kingdom narrowly avoided recession by mid-2023 with improved first-quarter results. This has been attributed to an economic boost from an influx of tourists and increased public spending during the coronation of King Charles III [6]. However, it is predicted that the UK economy will remain stagnant in the following quarters [7]. It is not clear whether a further recession will be avoided. Accurate early prediction of such economic crises would be of value, allowing action the be taken to avoid or minimise the effects of such recessions. Many organisations have attempted to model these economic fluctuations in the past. With the advent of machine learning, these algorithms have become increasingly refined, using multiple data sets and modelling algorithms.

Progression in machine learning algorithms has enabled us a large group of methods to test, for this paper we will use, Logistical Regression, linear discriminant analysis, K nearest neighbour, Decision tree classifier, gaussian naive Bayes, support vector classifier, neural network, random forest classifier, long short-term memory, convolutional neural network, and XGBoost.

In this study, we aim to show that accurately predicting recessions in the United Kingdom (a two-quarterly period of economic decline within the UK up to two quarters in advance), despite their differing underlying causes, is possible using machine learning techniques and diverse economic data. By modelling historic training data from the 1970's onwards we will assess the ability to predict the three very different recessions described above. In addition, this modelling will be tested on a second dataset, primarily composed of the 50-day simple moving average 10-year treasury rates minus 50-day simple moving average 3-month treasury rates. This US data has previously been effective at predicting recession a year in advance and will act as a control group [8].

Finally, we shall explore data that has not had results for whether there is a recession or not to see if the models can predict the future of economic recessions, by the time this paper is published most of the prediction dates will have been passed and thus we will be able to see whether or not these predictions were correct or not. Of the existing papers, reviewed below, that address recession prediction by machine learning algorithms, only a few incorporate multiple methods and none use multiple datasets to experiment on their validity. Our paper aims to address this shortcoming by exploring the use of multiple analytic methods with multiple datasets. The present paper is structured as follows: in Section 2 we describe briefly the relevant literature. Section 3 outlines the rational for dataset and algorithm selection and explains methodologies. The results are presented in section 4. In section 5 we discuss the implications of our findings and make suggestions for future research.

2. Literature Review

Cicceri et al attempted to predict a recession in Italy [8]. Fiscal parameters considered included GDP, Unemployment rate, Inflation, FTSE MIB (closed prices of the Italian main stock index), Italian treasury bill at issuing, ITA-treasury yield curve, Total Italian GDP and Debt, Italian GDP and Public Debt growth and Balance of Payments. Data was mainly gathered from the Eurostat website and the Bank of Italy. There was little data pre-processing or any feature engineering conducted. Analysis was undertaken using Autoregressive model (AR), Ordinary Least Squares regression (OLS), Nonlinear Autoregression models (NAR), Nonlinear Autoregressive with exogenous variables model (NARX), Support Vector Regression (SVR), K-Nearest Neighbours (KNN) and Boosted Trees (BT). The authors commented that whilst they used all of the above, NARX was most effective for this task, having the highest accuracy rate at 80%. In this publication, machine learning techniques provided more accurate results than OLS statistical models, supporting the validity of using machine learning to improve future economic predictions. They advocate the use of newer forms of machine learning, such as deep learning, to create more accurate predictions.

Puglia and Tucker applied machine learning methods to examine the use of financial market and macroeconomic variables to forecast US recessions [9]. Their data sets consisted of a monthly average of the 10-year Treasury spot yield, the 3-month log difference of end-of-month S&P 500 index values, Excess Bond Premium, the end-of-month values of the effective federal funds rate and the Federal Reserve Bank of Chicago's National Financial Conditions Index. They undertook no pre-processing beyond formatting the data to each algorithm. Random Forest, boosting (XGBoost and Light GBM), Support vector machine and Neural network algorithms were compared. They showed the Neural network algorithm had the highest accuracy at 0.867. They concluded that machine learning methods were able to capture important features of the joint empirical distribution of Treasury yields and other financial market and macroeconomic data over a recession indicator, that probit methods couldn't. In particular, machine learning methods, due to their flexibility, were able to capture the "non-linear" nature of those empirical distributions. The importance of using as much detail as possible when modelling was emphasised. This was contrary to Cicceri et al, but raised the interesting point that, if multiple methods are used in collaboration, a very accurate picture can be built up [8]. Puglia and Tucker stated that although machine learning will play a vital role in predicting non-linear events, the future of economic analysis will include other tools [9].

Nyman and Ormerod published results of a random forest machine learning method in future recession prediction and compared its results to an ordinary least squares regression [10]. Random forest was selected because of its ability to cope with noisy, non-linear, high-dimensional prediction problems. The data sets used in this paper were GDP, the three-month treasury bill rate, the yield on ten-year government bonds and the quarterly percentage change in the standard and poor five hundred index. They stated that these parameters were chosen based on theoretical links to GDP, but that these had not been validated. Apart from standard non-alteration methods required to allow data input, no data manipulation techniques were used before analysis. They demonstrated that the random forest algorithm was able to predict with reasonable accuracy a recession occurring in the next one and a half years. However, analysing the graphic results, the algorithms were generally late at setting the beginning and end points of a recession. It also struggled to predict the severity of the recession.

Döpke et al analysed a plethora of data consisting of 35 leading economic indicators [11]. The commoner parameters included U.S. effective federal funds rate, Money market rate, Discount rate, 3m-money market rate, and yields on debt. Most of this data was gathered from Deutsche Bundesbank, FRED, IFO Institute and OECD Monthly Economic Indicators. Apart from standard organisation, no data manipulation was undertaken. To conduct the analysis, they used boosted regression trees exclusively, stating the intention to use a proven machine learning algorithm to test diverse data sets on the effects they have on recessions. They conclude that using machine learning to predict recession should never replace traditional methods. However, using them in combination, proved reliable and improved predictive results. In keeping with Puglia and Tucker, the authors highlighted the importance of utilising diverse data sources and gave some guidance on the most important of these [9].

These publications try to predict future recessions using data specific to the one which has previously occurred. As a consequence, they proved only effective at predicting one type of recession [8,10]. With recent recessions being caused by very different and complex catalysts, it is clear that more than one factor must be considered when attempting to predict when a recession is likely. Multiple factors including the strength of the pound against other currencies, government bond prices, interest rates, inflation and unemployment figures must be considered. All of these are measurable and publicly available [11].

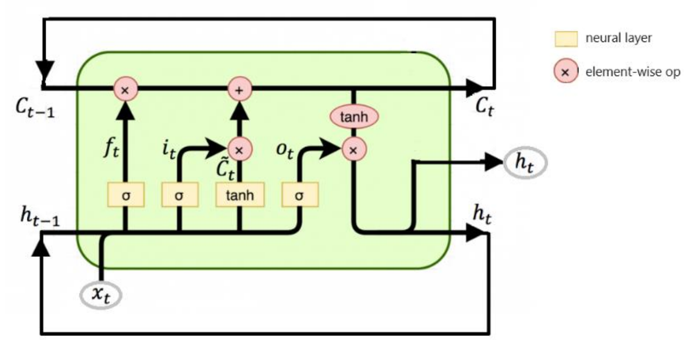
2.1. Machine Learning Methods

The different machine learning methods utilised in this project are discussed and compared. The advantages and disadvantages of each method and the statistics involved are highlighted. Long Short-Term Memory and XGBoost are described in detail.

2.1.1. Long Short-Term Memory (LTSM)

This is an advanced deep learning, recurrent neural network (RNN) that allows information to persist. It is a special type of RNN which resolves the vanishing gradient problem caused by traditional RNN and machine learning algorithms [12]. Its relative insensitivity to gap length is its advantage over other RNNs, hidden Markov models and other sequence learning methods. It aims to provide a short-term memory for RNN that can last thousands of timesteps, thus long short-term memory.

A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate (Figure 1). The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. Forget gates decide what information to discard from a previous state by assigning a previous state, compared to a current input, a value between 0 and 1. A (rounded) value of 1 means to keep the information, and a value of 0 means to discard it. Input gates decide which pieces of new information to store in the current state, using the same system as forget gates. Output gates control which pieces of information in the current state to output by assigning a value from 0 to 1 to the information, considering the previous and current states. Selectively outputting relevant information from the current state allows the LSTM network to maintain useful, long-term dependencies to make predictions, both in current and future time steps [13].



$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}H_{(t-1)} + b_{hf})$$

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{(t-1)} + b_{hi})$$

$$g_t = \tau\left(W_{ig}x_t + b_{ig} + W_{hg}h_{(t-1)} + b_{hg}\right)$$

$$o_t = \sigma\left(W_{io}x_t + b_{io} + W_{ho}h_{(t-1)} + b_{ho}\right)$$

$$c_t = f_t{}^\circ c_{(t-1)} + i_t{}^\circ g_t$$

Figure 1: Details of the LSTM algorithm, logic diagram and formulae. xt current input, h(t-1) previous output, c(t-1) previous state, ct new state, ft forget gate, it input gate, gt cell gate, ot output gate [14].

When training a traditional RNN using back-propagation, the long-term gradients which are back-propagated can "vanish" (that is, they can tend to zero) or "explode" (that is, they can tend to infinity), because of the computations involved in the process, which use finite-precision numbers. LSTM units partially solve the vanishing gradient problem, because LSTM units allow gradients to also flow unchanged. However, LSTM networks can still suffer from the exploding gradient problem [12]. The intuition behind the LSTM architecture is to create an additional module in a neural network that learns when to remember and when to forget pertinent information. In other words, the network effectively learns which information might be needed later on in a sequence and when that information is no longer needed [13]. LSTM is implemented in Python using the Keras library.

XGBoost (XGB): The eXtreme Gradient Boosting package is an open-source, advanced form of the random forest model, recommended for financial data by Puglia and Tucker [9]. It is an efficient and scalable implementation of a gradient-boosting framework, supporting various objective functions including regression, classification and ranking [15].

XGBoost operates on decision trees, models that construct a graph that examines the input under various 'if' statements. Whether the 'if' condition is satisfied influences the next 'if' condition and eventual prediction. The algorithm progressively adds more and more 'if' conditions to the decision tree to build a stronger model. XGBoost sets itself apart from other gradient-boosting techniques by using a second-order Taylor approximation of the scoring function. This approximation allows it to calculate the optimal 'if' condition and its impact on performance. This is represented graphically in Figure 2, with the generic algorithm underpinning the package, operating thus [16]:

Input: training set $\{(x_i, y_i)\}\frac{N}{i=1}$, a differentiable loss function $L(y, F(x))$, a number of weak learners M and a learning rate α.

1. Initialize model with a constant value: $f_{(0)}(x) = arg_\theta min \sum_{i=1}^{N} L(y_i, \theta)$
2. For m = 1 to M:

a. Compute the 'gradients' and 'hessians':

$$g_m(x_i) = \left[\frac{\partial L(y_i, f(x_{i)})}{\partial f(x_i)}\right]_{f(x)=f_{(m-1)}(x)} \qquad h_m(x_i) = \left[\frac{\partial^2 L(y_i, f(x_{i)})}{\partial f(x_{i)})^2}\right]_{f(x)=f_{(m-1)}(x)}$$

b. Fit a base learner (or weak learner) using the training set $\left\{x_{i,-}\frac{g_m(x_i)}{h_m(x_i)}\right\}_{i=1}^{N}$ by solving the optimization problem below:

c. $\emptyset_m = arg_{\emptyset\in\Phi} min \sum_{i=1}^{N} \frac{1}{2} h_m(x_{i)} \left[\emptyset(x_{i)} - \frac{g_m(x_{i)}}{h_m(x_{i)}}\right]^2$ *and* $f_m(x) = \alpha\emptyset_m(x)$

d. Update the model: $f_m(x) = f_{(m-1)}(x) + f_m(x)$

3. Output: $f(x) = f_{(M)}(x) = \sum_{m=0}^{M} f_m(x)$

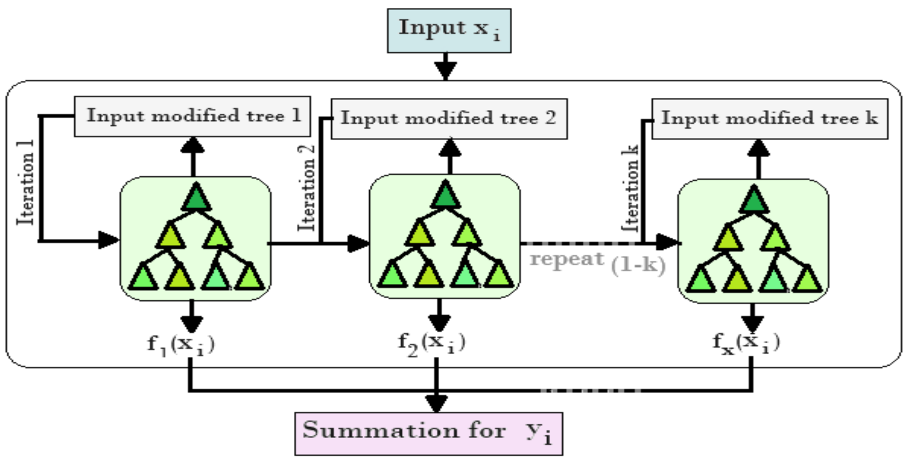


Figure 2: Simplified diagram of the XGBoost algorithm. During training, the regression tree function adds new trees that predict the residuals or errors of prior trees that are then combined with previous trees to make the final prediction [17].

Logistical Regression (LR): This is a machine learning method which extends the techniques of multiple regression analysis to research situations in which the outcome variable is categorical This makes it suited for datasets which have a binary identifier [18]. The algorithm works as follows:

$$Y = \alpha + \beta 1X1 + \beta 2X2 + \cdots + B\rho X\rho + \varepsilon = \alpha + \sum_{j-1}^{y} \beta jXj + \varepsilon$$

Where the Y-intercept (i.e., the expected value of Y when all X's are set to 0), β j is a multiple (partial) regression coefficient (i.e., the expected change in Y per unit change in Xj assuming all other X's are held constant) and ε is the error of prediction. If an error is omitted, the resulting model represents the expected, or predicted, value of Y. [18].

Linear Discriminant Analysis (LDA): Linear Discriminant Analysis is commonly used to identify the linear features that maximize the between-class separation of data, while minimizing the within-class scatter [19]. The algorithm works as followed:

$$\mathrm{Sw} = \frac{\mathrm{k\,i \in Ck\,(xi - mk)(xi - mk)T}}{N}$$

Where mk is the mean of kth class, and m is the mean of the data set [20].

K Nearest Neighbours (KNN): For a new input, the K nearest neighbours is calculated and the majority among the neighbouring data decides the classification for the new input. Even though this classifier is simple, the value of K plays an important role in classifying the unlabelled data. There are many ways to decide the values for K, but we can simply run the classifier multiple times with different values to see which value gives the most effective result [21]. This algorithm works as followed:

$$\sqrt{\sum_{i=1}^{n}(xi-yi)^2}$$

Find the K- nearest neighbours assign class containing the maximum number of nearest neighbours [21].

Decision Tree Classifier (DTC): A decision tree consists of a root node, several interior nodes, and several terminal nodes. The root node and interior nodes, referred to collectively as nonterminal nodes, are linked into decision stages, and the terminal nodes represent final classifications [22]. This algorithm works as followed:

$$E(T,X)=\sum_{c\in X}P(c)E(c)$$

Where T = Current state and X = Selected attribute.

Gaussian Naive Bayes (GNB): Bayesian classifiers, work based on the Bayesian rule and probability theorems. It has been proven that learning an optimal Bayesian classifier from training data is an NP-hard problem. A simplified version of the Bayesian classifier called naive Bayes uses two assumptions. The former is that, given the class label, attributes are conditionally independent and the latter is that, no latent attribute affects the label prediction process [23]. This works as followed:

$$\mathrm{p(x1,\ldots,xn|e)} = \prod_{i=1}^{n}p(xi|e)$$

Assume, the vector (x1, ..., xn) represents the n attributes of the instance x. Let c represent the class label of the instance x. The probability of observing x given the class label c can be computed by the previous equation [23].

Support Vector Classifier (SVC): We are looking for the optimal separating hyperplane between the two classes by maximizing the margin between the classes' closest points the points lying on the boundaries are called support vectors, and the middle of the margin is our optimal separating hyperplane [24]. The algorithm works as followed:

$$f(x)=sign\big(\big(W1,\theta(x)\big)+b\big)$$

It can be shown that the optimal, in terms of classification performance, hyper-plane is the one with the maximal margin of separation between the two classes [24].

Neural Network (NN): A feed-forward neural network can be regarded as a nonlinear mathematical function which transforms a set of input variables into a set of output variables. The precise form of the transformation is governed by a set of parameters called weights whose values can be determined based on a set of examples of the required mapping [25].

$$a=\sum_{j-1}^{d}WiXi+W0$$

Where the offset parameter W is called a bias (and corresponds to the tiring threshold in a biological neuron). Formally, the bias can be regarded as a special case of a weight from an extra input whose value x0 is permanently set to + 1 [25].

Random Forest Classifier (RFC): The random forest classifier consists of a combination of tree classifiers where each classifier is generated using a random vector sampled independently from the input vector, and each tree casts a unit vote for the most popular class to classify an input vector [26].

$$\sum\sum_{j\neq i}(f(Ci,T)/|T|)(f(Cj,T)/|T|)$$

Where f (C_i, T) / | T | is the probability that the selected case belongs to class Ci [11].

Convolutional Neural Network (CNN): As neural network models have shown promise, it was decided to test a convolutional neural network (Figure 3). Although generally used with image data, it may still prove useful. It is a well-known deep-learning architecture inspired by the natural visual perception mechanism of living creatures. It can obtain effective representations of the original image, which makes it possible to recognize visual patterns directly from raw pixels with little to no pre-processing [27].

$$Convulution: W_{out} = \frac{W - F + 2P}{s} + 1$$

$$ReIU: f(k) = \max(O, k)$$

$$Softmax: Softmax(X_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$$

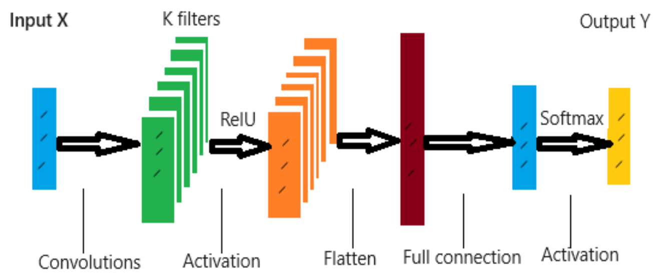


Figure 3: Diagram illustrating the basic algorithm of CNN. Inputs are mapped during convolution, further modified and the matrix flattened, before passing to the full connection neural network and ultimately outputted [28].

3. Materials and Methodology

Multiple machine learning and neural network methods were trained using a broad spectrum of historical financial data from 1900 onwards. Their ability to accurately predict known resections until the present time was investigated. Confusion matrices, average F1, recall and accuracy for 100 iterations of each model were used to assess overall performance against actual events. Recent data was then used to produce a recession prediction from 2023 to 2024 for each model. A detailed description and rationale for the inclusion of the UK and USA data parameters is presented. The various machine learning methods employed are described and their analysis methods are explained. Based on available results the more advanced algorithms, Long Short-Term Memory and XGBoost showed the most potential and are described in greater detail.

3.1. Dataset Selection

We used data such as GDP and Unemployment Rate. The aim was to include a breadth of financial, economic and sociological parameters to maximise the predictive power for the varied types of recession experienced in recent times.

• **United Kingdom GDP:** Gross domestic product or GDP is a measure of the size and health of a country's economy, based on the size of goods and services produced over some time (usually one quarter or one year) [29].

• **Unemployment Rate:** The proportion of the labour force that is unemployed. This includes people of working age who are without work, including those economically inactive. It excludes those classed as long-term sick or disabled [30].

• **Inflation:** The rate of change in prices for goods and services over time. Measures of inflation and prices include the Consumer Price Index (standardised basket of goods and services consumed by households), Producer Price Index and the House Price Index.

• **FTSE 100 Index:** This is the main stock index in the UK. The way this is measured is by weighing all stocks listed on the London Stock Exchange by market capitalisation. First published in 1984 so lacks long-term historical data.

• **United Kingdom Yield Curve:** The percentage return from a set investment in United Kingdom government bonds. Bonds have a fixed price and rate of interest set by the government, so values can be easily calculated.

• **United Kingdom Debt:** The total sum of money owed or due by the central government. The total government debt is simply the accumulation of all the previous years' deficits.

• **USA data - 50-day simple moving average 10-year treasury rates minus 50-day simple moving average 3-month treasury rates:** A large volume of detailed historic economic data is readily available for the USA. This parameter has predicted all recent USA recessions a year in advance. The inclusion of this data gives an international perspective and serves to act as both a control and benchmark for the current project. The yield curve depicts the interest rates of treasury securities of various maturities that have equal credit quality and the same risk characteristics [31]. Subtracting the 50-day simple moving average 10-year treasury rates from the 50-day simple moving average 3-month treasury rates produces a graph which forms a clear dip or trough one year before a recession. As a leading indicator of recession prediction, economists typically incorporate the yield spread in probit models to forecast the probability of a recession [31]. This dataset can be found at https://fred.stlouisfed.org/series/T10Y3M.

• **Pre-processing:** As with previous publications, no attempt to adapt, manipulate or pre-analyse the data sets was undertaken. Only standard machine learning pre-processing, such as adding the in our case column and making sure there are no gaps or erroneous data was performed to ensure the algorithms ran. An example of this in the American dataset occurred with data recorded daily except on Sundays and bank holidays.

3.2. Outcome Measures

• **Accuracy:** The ratio between the number of correctly classified samples and the overall number of samples [32,33]. This can be calculated by dividing the number of correct predictions by the total number of predictions:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

This gives a basic measure of how well each algorithm is performing. Given the nature of the project, a high level of accuracy would be expected. Therefore, an accuracy of greater than 70% for a majority of the algorithms, will be used to define success.

• **Recall:** defined as the percentage of data samples that a machine learning model correctly identifies as belonging to a class of interest (the "positive class") out of the total samples for that class [34]. This can be calculated by dividing the number of positive samples classified correctly as positive by the total number of positive samples [32]:

$$Recall = \frac{TP}{TP + FP}$$

This allows a measure of how many of the few periods of recession the algorithms correctly identified. A recall of above 50% for a majority of algorithms will be considered a success. The value is set relatively low because where multiple parameters in a row predict a recession, it matters less that all are around 50:50.

• **F1 Score:** This is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores and provides a more realistic view of the model as a whole [34]. It "Fβ-Measure is a trade-off between PPV and TPR" [36]. This can be calculated by dividing precision multiplied by recall by precision plus recall and then multiplying the results by two:

$$F1 = 2 * \frac{\frac{TP}{TP + FN} * recall}{\frac{TP}{TP + FN} + recall}$$

Since new data will be constantly fed into the algorithms, a high F1 score is not required. If, for example, it predicts correctly half of the time that it is true, then ten out of twenty true results in a tight period would allow an accurate determination that it is probably true.

4. Results

The ability of each of the models to correctly predict known recessions until late 2022 from historical economic data is presented. Training results for recall, accuracy and F1, averaged over 100 iterations of each model are shown in Table 1 for the UK dataset and Table 2 for the USA dataset. Overall, accuracy was high for all models for both the UK and USA datasets, with values between 1.0 and 0.81. Recall and F1 showed greater variation between models. For both the UK and USA datasets KNN, DTC, RFC, LSTM and XGB returned noticeably higher values than the other models, with recall between 1.0 to 0.75 and F1 between 1.0 to 0.69. Based on these parameters, LSTM and XGB performed particularly well with recall over 0.99 and F1 over 0.97 for the UK dataset. SVC and CNN had a recall and F1 of zero, suggesting an inability to predict positive results in this study. LDA also performed poorly when tested with the USA dataset, with a recall of 0.07 and F1 of 0.12. The performance for the remaining models lay between these extremes with recall from 0.24 to 0.54 and F1 0.32 to 0.49.

4.1. Results Presentation

Model	Recall	Accuracy	F1
LR	0.238	0.900	0.343
LDA	0.376	0.898	0.444
KNN	0.750	0.961	0.803
DTC	0.784	0.953	0.786
GNB	0.433	0.810	0.317
SVC	0	0.892	0
NN	0.365	0.918	0.489
RFC	0.722	0.962	0.795
LSTM	1.000	0.994	0.968
CNN	0	0.908	0
XGB	0.995	0.988	0.998

Table 1: Average results for 100 iterations of each model using the United Kingdom's dataset

Model	Recall	Accuracy	F1
LR	0.343	0.889	0.424
LDA	0.071	0.871	0.118
KNN	0.715	0.920	0.686
DTC	0.771	0.931	0.727
GNB	0.536	0.870	0.498
SVC	0	0.879	0
NN	0.326	0.886	0.407
RFC	0.770	0.929	0.725
LSTM	0.962	0.990	0.955
CNN	0	0.910	0
XGB	0.746	0.967	1.000

Table 2: Average results for 100 iterations of each model using the USA's Dataset

The corresponding confusion matrix results from running each algorithm 100 times on the UK and USA datasets are shown in Tables 3 and 4. In keeping with the observations on recall, accuracy and F1, KNN, DTC, RFC, LSTM and XGB performed noticeably better, correctly identifying 2 out of 3 relatively uncommon Positive recession events. Against the USA dataset, LSTM and XGB identified 11 out of 12 Positive events. SVC and CNN failed to identify any Positive recession events. LDA also performed poorly when tested with the USA dataset, identifying only 1 in 12 Positive events. The remaining models correctly identified 1 in 3 of the Positive events for both datasets.

Overall False return rates for KNN, DTC, RFC, LSTM and XGB models were between 3% and 5% for the UK data and 1% to 7% for the USA data. This compared favourably with values between 8% and 13% for most of the remaining models. The GNB model performed worst with 19% False returns. False-positive rates were low for all models at under 4%. The GNB model was an exception with a 13% False-Positive rate for the UK dataset. The inability of SVC and CNN to identify Positive events was reflected in the higher False-Negative values of 11% and 12% respectively. False-Negative rates were noticeably lower for the better-performing KNN, DTC, RFC, LSTM and XGB models at under 4%. LSTM and XGB performed particularly well with the USA dataset, will False-Negative values under 1%.

Model	True-Pos	True-Neg	False-Pos	False-Neg
LR	3%	87%	1%	9%
LDA	4%	86%	4%	6%
KNN	8%	88%	1%	3%
DTC	8%	87%	2%	3%
GNB	4%	77%	13%	6%
SVC	0%	89%	0%	11%
NN	4%	88%	1%	7%
RFC	8%	88%	<1%	3%
LSTM	8%	88%	1%	4%
CNN	0%	88%	0%	12%
XGB	8%	89%	0%	3%

Table 3: Confusion matrix results from running each algorithm 100 times on the UK dataset. The actual dataset consisted of 11% positive and 89% negative events.

Model	True-Pos	True-Neg	False-Pos	False-Neg
LR	4%	85%	3%	8%
LDA	1%	86%	2%	11%
KNN	9%	84%	4%	3%
DTC	9%	84%	4%	3%
GNB	7%	81%	7%	5%
SVC	0%	88%	0%	12%
NN	4%	85%	3%	8%
RFC	9%	85%	3%	3 %
LSTM	11%	88%	1%	1%
CNN	0%	89%	0%	11%
XGB	11%	89%	<1%	<1%

Table 4: Confusion matrix results from running each algorithm 100 times on the USA dataset. The actual dataset consisted of 12% positive and 88% negative events.

4.2. Future Recession Prediction

The USA dataset for the previous 12 months was run with the machine learning models that had been trained as detailed above. These provided recession predictions one year in advance, for the period May 23 to April 24 as shown in Table 5. SVN and CNN were excluded from the analysis due to their earlier failure to return positive predictions. Six of the nine models (KNN, DTC, GNB, NN, RFC and XGB) predicted the USA avoiding recession during this period. However, the predicted probability did increase from 0 on May 22 up to a maximum of 0.54 by November 22, before rapidly reducing once more. XGB predicted a low rate of 0.01 throughout. LR and LDA initially predicted no recession, but the probability then rapidly increased during the year to 0.97 and 0.83 respectively by April 2023. LSTM predicted recession throughout, with the risk rising to 0.90 by April 2024. Although recession may be avoided, most models suggest an increasing risk.

Data time	Predicted Probability of Recession within 1yr								
period	LR	LDA	KNN	DTC	GNB	NN	RFC	LSTM	XGB
May 22	0.01	0.03	0	0	0	0	0	0.29	0.01
Jun 22	0.01	0.04	0	0	0	0	0	0.52	0.01
Jul 22	0.14	0.16	0.12	0.11	0.20	0.19	0.12	0.58	0.01
Aug 22	0.30	0.27	0.24	0.23	0.50	0.45	0.23	0.62	0.01

Sept 22	0.27	0.26	0.37	0.39	0.47	0.44	0.40	0.65	0.01
Oct 22	0.27	0.25	0.18	0.19	0.43	0.38	0.21	0.67	0.01
Nov 22	0.58	0.44	0.11	0.13	0.54	0.40	0.13	0.69	0.01
Dec 22	0.83	0.62	0	0	0.25	0.14	0	0.71	0.01
Jan 23	0.87	0.69	0.08	0.08	0.16	0.12	0.08	0.73	0.01
Feb 23	0.92	0.74	0	0	0.05	0.02	0	0.74	0.01
Mar 23	0.91	0.73	0	0	0.07	0.03	0	0.76	0.01
April 23	0.97	0.83	0	0	0	0	0	0.90	0.01

Table 5: Predicted probability of recession in one year (May 23 to April 24) using the USA dataset between May 22 and April 23 for each model. SVN and CNN were excluded due to never returning positive predictions.

5. Discussion

The task of recession prediction, requiring the accurate identification of relatively uncommon true events from a mass of false events, represents a challenge for any machine learning algorithm. In this study, a wide variety of differing models were tested to assess their suitability. All showed a consistently high level of accuracy using the training data sets. However, recall and F1 varied significantly. In the current study, CNN and SVC proved very inadequate with both performance measures between 0.1 and 0. This was reflected in the confusion matrix results with the model's incapable of returning any true predictions. LDA showed similar inadequate results for the USA dataset. LR, GNB and NN were more reliable with recall and F1 values between 0.25 and 0.55. However, these models only reported 1 in 3 true-positive events and had an excess of false reporting. KNN, DTC and RFC performed much better. With recall and F1 between 0.7 and 0.8, they correctly identified 2 out of 3 true-positive events, with fewer false reports. LSTM, and XGB performed best overall, with recall and F1 between 1.0 and 0.75. They returned consistently accurate predictions reaching 11 out of 12 true-positive rates against the USA data.

Overall, there was little difference in performance between the UK and USA datasets for most models. LDA was the principal exception, with significantly poorer performance for the USA data. Given that the data sets were specifically chosen as representative indicators of economic conditions in each country and recessions are generally global, this finding of conformity between UK and USA data was to be expected. Based on the present findings, LDA, SVC, NN, GNB and CNN cannot be recommended for economic recession prediction under these conditions. The other models investigated were more suitable, with LSTM and XGB delivering a superior performance. This observation that the newer and more complex neural networks have great potential in the economic world has also been made by previous authors.

Cicero et al reported an Italian study investigating multiple models, with methods similar to the current work [8]. They found that the Nonlinear Autoregressive with Exogenous Variables (NARX) model of the neural network outperformed earlier algorithms with an 80% accuracy of prediction. Nyman and Ormerod reported strong recession predictions from both UK and USA data using RTC [10]. Although the timing of onset was often late, it gave no false predictions. Döpke et al reported positive predictions using boosted regression trees (BRT) model [11]. They did however state that such techniques were currently unable to provide a full economic picture and should be used in conjunction with traditional forecasting methods. Puglia and Tucker compared neural networks to other predictive methods using USA data [9]. Contrary to other publications, they found that these did not perform markedly better than other prediction methods. This was despite the neural networks used having an accuracy of 0.87. They found that, depending on the test methodology used, they were outranked by probit and tree regression techniques.

Since recessions occur infrequently, they constitute only a tiny fraction of total economic data inputted. As with previous studies, to improve each model's ability to identify these events, duplicate positive rows were included in the training data used in the current study. Although this undoubtedly served to improve recall, it increases the risk of the model overfitting when used for future predictions. Puglia and Tucker expressed similar concerns when devising training for low-frequency events [9]. This issue of the varying interaction of data parameters affecting model outputs is raised by Döpke et al [11]. They emphasise the importance of data set selection to achieve success. The inclusion of a broad spectrum of representative social and economic parameters is needed to maximise the identification recessions with widely differing underlying causes.

There was a noticeable variation in the prediction of a future recession (during 2023-24) returned by each model in the current study. Except for XGB, all models predicted a low initial risk, which rose through the first half of the period. Thereafter modelling diverged, with most reporting a falling subsequent risk. However, LR, LDA and LSTM showed this continuing to rise towards near certainty. This seems counter-intuitive for LR and LDA, as these two models had the lowest recall and F1 values. Overall, it would appear that, although recession may be avoided, most models do suggest an increasing risk. The true situation and the accuracy of these predictions will become apparent by early 2024.

These discrepancies may reflect economic conditions unseen in the training data. During the period 1990 to 2020, used for training the models, individual economic parameters remained relatively stable, showing only modest variations between recessions. From late 2021 onwards, including the period of future recession prediction, there were sudden and significant variations in many of the datasets. For example, inflation rose sharply to 10% and GDP fell by 11%, before rapidly rebounding [37]. During the training period the USA '10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity' remained positive until mid-2022. It then dipped sharply, becoming very negative throughout the prediction period [38]. Consequently, models may be displaying overfitting or an inability to correctly identify or factor in these unprecedentedly low values.

The principle of using machine learning methods in recession prediction up to one year in advance (using data from the USA) is confirmed in the current study. However, the performance of the individual algorithms varied, with several proving poor at accurately identifying the rare recession events amongst abundant non-recession events. Some of the possible factors leading to this finding have been discussed. The more advanced neural network models, especially LSTM proved the most accurate. To further improve performance, research into which of the recently developed advanced neural network models best suits the task, is required. These need optimising to accurately predict infrequent recession events and minimise false reporting. Further refinement of the data set used for model training, to improve data relevance and allow a lower number of duplications during training, could aid reliability. The creation of a dataset and algorithm capable of processing global rather than country-specific data and events has potential. This has validity in an interconnected global economy where it is rare for a recession to affect only one country [39,40].

Acknowledgements

Mr Arion Thomson, a Fellow Undergraduate, for his assistance with the USA dataset.

Federal Reserve Bank of St. Louis, for the USA dataset.

References

1. Robeck, D. (2023). What is a recession?.
2. Bianco, K. M. (2008). The subprime lending crisis: Causes and effects of the mortgage meltdown (pp. 1-21). New York: CCH, Wolters Kluwer Law & Business.
3. Grinin, L., & Korotayev, A. (2020). COVID-19 pandemic, geopolitics, and recession. International Center for Education and Social and Humanitarian Studies. Working Paper, 4.
4. HM Treasury. (2022). The Growth Plan. UK Government.
5. United Nations. (2023). One year of the war in Ukraine leaves lasting scars on the global economy. UN Department of Social Affairs.
6. Harmsworth, E. (2023). Burden or benefit? The economics of King Charles' coronation. The Economic Times, May 2023.
7. CBI. (2023). UK economy set to grow and business investment to rise following brush with recession – CBI Economic Forecast.
8. Cicceri, G., Inserra, G., & Limosani, M. (2020). A machine learning approach to forecast economic recessions—an italian case study. Mathematics, 8(2), 241.
9. Puglia, M., & Tucker, A. (2020). Machine learning, the treasury yield curve and recession forecasting.
10. Nyman, R., & Ormerod, P. (2017). Predicting economic recessions using machine learning algorithms. arXiv preprint arXiv:1701.01428.
11. Döpke, J., Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted regression trees. International Journal of Forecasting, 33(4), 745-759.
12. Hochreiter, S. and Schmidhuber, J., (1997). Long Short-Term Memory. Neural Comput, 9 (8): p1735–1780.
13. Wikipedia contributors. (2023). Long short-term memory. In Wikipedia, The Free Encyclopedia.
14. McCaffrey, J. (2018). Understanding LSTM Cells Using C#. MSDN Magazine, 33(4), p58-61.
15. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.
16. Wikipedia contributors. (2023). XGBoost.
17. Zou, M., Jiang, W., Qin, Q., Liu, Y., Li, M. (2022). Optimized XGBoost Model with Small Dataset for Predicting Relative Density of Ti-6Al-4V Parts Manufactured by Selective Laser Melting. Materials. 15(15): p5298.
18. Dayton, C. M. (1992). Logistic regression analysis. Stat, 474, 574.
19. Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford university press.
20. Ioffe, S. (2006). Probabilistic linear discriminant analysis. In Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part IV 9 (pp. 531-542). Springer Berlin Heidelberg.
21. Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019, May). A brief review of nearest neighbor algorithm for learning and classification. In 2019 international conference on intelligent computing and control systems (ICCS) (pp. 1255-1260). IEEE.
22. Swain, P. H., & Hauska, H. (1977). The decision tree classifier: Design and potential. IEEE Transactions on Geoscience Electronics, 15(3), 142-147.
23. Jahromi, A. H., & Taheri, M. (2017, October). A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. In 2017 Artificial intelligence and signal processing conference (AISP) (pp. 209-212). IEEE.
24. Meyer, D., & Wien, F. T. (2015). Support vector machines. The Interface to libsvm in package e1071, 28(20), 597.
25. Bishop, C. M. (1994). Neural networks and their applications. Review of scientific instruments, 65(6), 1803-1832.
26. Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.
27. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). Recent advances in convolutional neural networks. Pattern recognition, 77, 354-377.
28. Jin, X. B., Yang, N. X., Wang, X. Y., Bai, Y. T., Su, T. L., & Kong, J. L. (2020). Deep hybrid model based on EMD with

classification by frequency characteristics for long-term air quality prediction. Mathematics, 8(2), 214.
29. Bank of England, (2019). What is GDP? Bank of England.
30. OECD (2022), OECD Employment Outlook 2022: Building Back More Inclusive Labour Markets, OECD Publishing, Paris.
31. Joshi, S. (2020). Forecasting the Leading Indicator of a Recession: The 10-Year minus 3-Month Treasury Yield Spread. arXiv preprint arXiv:2009.05507.
32. Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics, 21(1), 1-13.
33. Google Developers, (2022). Classification: Accuracy | Machine Learning Crash Course.
34. Iguazio, (2022). What is recall in Machine Learning? Iguazio.
35. Kundu, R. (2023). F1 Score in Machine Learning: Intro & Calculation.
36. Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC medical imaging, 15(1), 1-28.
37. ONS (2023), Gross Domestic Product, Inflation & prices indices.
38. FRED Economic Data, (2023). 10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Ma-turity (T10Y3M).
39. Design Time, (2022). Series NARX Feedback Neural Networks - MATLAB &amp, Simulink. MathWorks.
40. Getting started, (2023). A guide to scikit learn.