Application of Supervised Machine Learning Techniques to Forecast the COVID-19 U.S. Recession and Stock Market Crash

Rama K. Malladi

Accepted: 8 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Machine learning (ML), a transformational technology, has been successfully applied to forecasting events down the road. This paper demonstrates that supervised ML techniques can be used in recession and stock market crash (more than 20% drawdown) forecasting. After learning from strictly past monthly data, ML algorithms detected the Covid-19 recession by December 2019, six months before the official NBER announcement. Moreover, ML algorithms foresaw the March 2020 S&P500 crash two months before it happened. The current labor market and housing are harbingers of a future U.S. recession (in 3 months). Financial factors have a bigger role to play in stock market crashes than economic factors. The labor market appears as a top-two feature in predicting both recessions and crashes. ML algorithms detect that the U.S. exited recession before December 2020, even though the official NBER announcement has not yet been made. They also do not anticipate a U.S. stock market crash before March 2021. ML methods have three times higher false discovery rates of recessions compared to crashes.

Keywords Machine learning · Forecasting · Financial econometrics · Recession · Stock market crash

JEL Classification G11 · G17 · C01 · C5 · C58 · C63

Rama K. Malladi is with California State University, Dominguez Hills, in California. He received CFA Charter in 2006 and Oracle certification in database administration in 1999. This paper has benefited from participants’ feedback on the Machine Learning courses and labs organized by the EDHEC Business School in Nice, France. The feedback received in the “Research with Machine Learning Applications” session of the 96th annual conference (2021) of the Western Economic Association helped refine the paper’s narrative.

* Rama K. Malladi
rmalladi@csudh.edu

1 California State University Dominguez Hills, Carson, CA, USA
1 Introduction

Machine Learning (ML) is a field that develops algorithms designed to be applied to datasets, with the main areas of focus being prediction, classification, and clustering (Athey, 2018). ML is broadly recognized as a subset of Artificial Intelligence (AI). AI is a general field that encompasses machine learning and deep learning, but that also includes many more approaches that do not involve any learning. Early chess programs, for instance, only involved hard-coded rules crafted by programmers and did not qualify as machine learning (Chollet & Allaire, 2018).

The industry’s adoption of AI/ML is rapid—The 2020 McKinsey Global Survey on AI suggests that organizations use AI to generate value. Increasingly, that value is coming in the form of revenues. A small contingent of respondents from various industries attribute 20 percent or more of their organizations’ earnings before interest and taxes (EBIT) to AI. These companies plan to invest even more in AI in response to the COVID-19 pandemic and its acceleration of all things digital (McKinsey & Company, 2020). ML’s emergence as a mainstream tool is relatively recent (Pyle & San José, 2015).

In traditional programming, humans input plausible rules using a programming language (code) and test those rules on a dataset to derive answers. Rules that do not improve the validity of answers are discarded. A plethora of rules, such as multiple regression, have been invented over the last few decades. Intuition is required to explain why certain rules work and others do not. On the other hand, a machine-learning system is trained rather than explicitly programmed (Chollet & Allaire, 2018). Intuition fails in higher dimensions (Domingos, 2012).

ML does not follow the decades-old paradigm of humans providing rules. Instead, ML expects humans to provide answers and datasets, and the machine finds rules. These machine-found rules can be applied to new data to predict future outcomes. As more data is fed, ML provides more answers. Unsupervised ML involves finding clusters of similar observations in their co-variates; it is commonly used for video, images, and text (Athey, 2018).

In the case of supervised machine learning, humans classify these predicted outcomes as right or wrong. The machine learns from the wrongly-predicted answers and adjusts the rules to avoid wrong predictions. It may be impossible to intuitively explain the machine-found rules, such as Support Vector Machine (SVM) or AdaBoost. Traditional econometricians face two main hurdles in adopting ML: (1) Moving from older intuitive techniques such as ordinary least squares (OLS) to newer non-intuitive methods, such as AdaBoost, often implemented in a hard-to-learn Python programming language; (2) Letting machines train themselves and take over economic and financial forecasting from humans through a black-box approach. The latter is a bigger hurdle than the former.

Data analysis in statistics and econometrics can be broken down into four categories: (1) prediction, (2) summarization, (3) estimation, and (4) hypothesis testing. Machine learning is concerned primarily with prediction (Varian, 2014). Recession prediction is a popular theme for scholars in economics: Google Scholar shows more
than 200,000 publications on this topic. Similarly, the stock market prediction is a more popular theme for financial market researchers: Google Scholar yields more than 1.8 million results. Numerous determinants of recessions and stock markets were found by researchers all over the world. However, many of them fail to forecast recessions and stock market crashes. The often-quoted example of the Queen of the U.K. questioning academics at the London School of Economics on why they did not see the 2008 financial crisis coming illustrates the point (Palmer, 2008; Pierce, 2008).

This paper’s main contribution is to forecast recessions and stock market crashes using the new supervised ML methods in a walk-forward out-of-sample setting. NBER definition, obtained from the Business Cycle Dating website, is used to classify “recession” months. Stock market (as measured by the S&P 500 index) peak to trough percent drop (also called the “maximum drawdown”) is used to classify “crash” months. A 20% or more drawdown is classified as a stock market crash. Some people also call this a “bear market.”

In particular, the following four research questions are addressed: could the recent supervised ML techniques have predicted a COVID-19 recession and a stock market crash in the U.S. in March 2020? Is the 2020 recession over, or is it continuing? Will the stock market crash again in the next three months (i.e., before March 2021)?

The rest of the paper is organized as follows: the next section reviews the evolving ML literature with a backdrop of recessions and stock market crashes. The section on data explains 134 macroeconomic variables in this study. The methodology section briefly describes the ML methods deployed in this paper, followed by the results section. The final section presents conclusions and identifies the scope for further research.

2 The Literature on Machine Learning, Recessions, and Crashes

Alan Turing, a British mathematician, questioned, “Can machines think?” in his 1950 article in the Mind journal (Turing, 1950). Many consider Turing as the beginning of AI. Pinar Saygin et al. (2000) review the 50-years of the Turing Test, philosophical debates, practical developments, and repercussions in related disciplines. In 1959, Arthur Samuel, a computer gaming and artificial intelligence expert, coined the term ‘Machine Learning’ (Samuel, 1959). The Samuel Checkers-playing Program appears to be the world’s first self-learning program and a very early demonstration of a fundamental concept of AI (Professor Arthur Samuel, 1990). AI is concerned with understanding and building intelligent entities—machines that can compute how to act effectively and safely in various novel situations (Russel & Norvig, 2013).
The academic literature on machine learning is broad and deep, with more than 4.9 million results on Google Scholar. It is one of the fastest-growing research areas. However, the lens used for academic research in traditional economics is not essential to the machine learning community—traditional econometricians focus on causal effects and identification, whereas machine learners focus on prediction. As a result, the adoption of AI is slow in economics. However, the recent focus on AI and ML is changing the picture rapidly. Events such as the Economics of AI conference and continuing education webcasts of the American Economic Association have increased the visibility of ML in the economics area. Many now believe that ML will have a dramatic impact on the field of economics within a short time frame (Athey, 2018).

The main difference in approach between ML and traditional econometrics is in model selection. In traditional econometrics, a researcher, based on experience and intuition, picks a model and estimates parameters. In contrast, ML methods use “tuning” (produce best out of sample prediction accuracy) and “k-fold cross-validation” (divide data into k equal bins or folds) to select the best model. The ML procedure works because prediction quality is observable: both predictions \( \hat{y} \) and actuals \( y \) are observed. Contrast this with parameter estimation, where typically, we must rely on assumptions about the data-generating process to ensure consistency. Observability alone would not make prediction easier since the algorithm still needs to sort through an extensive class of functions. ML uses “regularization” to penalize models for excessive complexity since simpler models tend to work better for out-of-sample forecasts (Athey, 2018; Hastie et al., 2009; Mullainathan & Spiess, 2017; Varian, 2014).

The finance area has adopted ML from the beginning (Dixon et al., 2020; Leippold et al., 2021; Prado, 2018). The search for alpha through information mining, fraud detection and prevention, loan approvals, default prediction, and sentiment analysis of alternative data are reasons behind this adoption. Alternative data is defined as information outside the usual scope of securities pricing, company fundamentals, or macroeconomic indicators (Dixon et al., 2020).

For specific applications of ML in finance, refer to the ML literature on credit-risk modeling (Khandani et al., 2010), credit card fraud detection (Adewumi & Akinyelu, 2017), corporate fraud detection (Hajek & Henriques, 2017; Perols, 2011), FinTech (Philippon, 2016), asset management (Prado, 2020), daily stock returns (Zhong & Enke, 2019). However, publications have not yet appeared to apply ML to validate the U.S. recession and the stock market crash following the COVID-19 global health crisis (GHC). It is a matter of time before numerous papers appear on this topic, as this area is ripe for ML methods.

Most-widely accepted definition of U.S. recession comes from the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER), which officially dates the beginnings and ends of U.S. recessions. The NBER’s traditional
definition emphasizes that a recession involves a significant decline in economic activity spread across the economy and lasts more than a few months. Since recessions have a significant negative impact on people’s lives with rising unemployment levels and potential loss of wealth, understanding and forecasting recessions is a critical task.

Many analysts use specialized models to predict the onset of a recession. These models are useful primarily because the economy’s behavior during periods of transition between expansion and recession is fundamentally different from when a recession is not imminent (Filardo, 1999; Hymans et al., 1973). Five such popular business cycle models are: (1) simple rules-of-thumb using ten of the Conference Board’s composite index of leading indicators (CLI); (2) probability model of imminent recession using the CLI and recursive equations (Neftiçi, 1982); (3) Probit or probability of recession model based on four variables; (4) GDP forecasting model based on a four-variable vector autoregression (VAR); (5) Stock-Watson recession prediction model to predict probabilities of a recession using 45 variables (Stock & Watson, 1992). Although all the models provide helpful information about imminent recessions, none of them was foolproof. The ex-post analysis showed that some models, such as the CLI rules-of-thumb and Neftiçi model, were unreliable. Thus, leaving three reliable models: the probit model, GDP forecasting model Stock-Watson model (Filardo, 1999).

Other researchers continued to refine the recession predictability models. They found a plethora of recession predictability factors such as consumer sentiment (Howrey, 2001), yield spread (Moneta, 2005), data revisions based on the Real-Time Data Set for Macroeconomists (RTDSM)—a publicly-available macroeconomic dataset at Philadelphia Fed5 (Croushore & Stark, 2001), interest-rate spread (Ahrens, 2002; Kauppi & Saikkonen, 2008), shocks to aggregate demand and technology (Shapiro & Watson, 1988), household deleveraging (Glick & Lansing, 2009), problems in housing and consumer durables (Leamer, 2008), and the great moderation—a result of smarter counter-cyclical monetary policy (Bernanke, 2012; Blanchard & Simon, 2001).

One major issue plaguing the recession prediction researchers has been the time-varying relevance of various leading indicators (Qi, 2001) and the inability to capture any nonlinear relationships in the data (Zhang, 2004). Traditional econometric models struggle to estimate model parameters when the underlying regimes change, and the deterministic factors constantly shift in importance. The key strength of ML lies precisely in the area where the traditional econometric models are weak. Machines can model complex and high-dimensional data generation processes, sweep through millions of model configurations, and robustly evaluate and correct the models in response to new information (Dhar, 2013; Dixon et al., 2020).

Financial market forecasting is the holy grail of forecasters. The desire to forecast stock prices has been eternal (Cowles, 1933; Dokko & Edelstein, 1989). Almost every technique and factor under the sun has been (and will be) tried to forecast stock returns. There are tremendous monetary gains for accurate forecasters. Since

---

5 RTDSM dataset at Philadelphia Fed:
the factor list is endless, the less said, the better about the factors behind stock market returns. If compelled to provide some, the author’s favorites are planetary movements (Sivakumar & Sathyanarayanan, 2007), moon phases (Borowski, 2015), and unlucky listing codes (Bai et al., 2020; Hirshleifer et al., 2016).

The characterization of large stock market moves, especially large negative price drops, is of profound importance for risk management and portfolio allocation (Johansen & Sornette, 2010; Sornette, 2017). Galbraith’s classic book (Galbraith, 1961) still provides the most commonly accepted explanation of the 1929 boom and crash. Galbraith emphasizes the irrational element—the mania—that induced the public to invest in the bull market. This eagerness to buy stocks was then fueled by an expansion of credit in the form of brokers’ loans that encouraged investors to become dangerously leveraged (White, 1990).

The causes of what halts and reverses the direction of the over-heated market are not as certain. Many arguments such as culture shift (Lowenstein, 2004), intrinsic bubbles (Froot & Obstfeld, 1989), irrational exuberance (Shiller, 2015), leverage cycle (Geanakoplos, 2010), liquidity (Amihud et al., 1990), jumps (Eraker et al., 2003; Santa-Clara & Yan, 2010), and fear and greed (Shefrin, 2002) are some of the many plausible explanations.

It is intriguing to see how well ML performs when it joins this race to predict recessions and stock market crashes. In particular, could the supervised ML techniques have predicted the Covid-19 recession and the ensuing 30% U.S. stock market crash in March 2020 during the COVID-19 pandemic? More importantly, will the stock market crash again in the next three months (i.e., before March 2021)?

| State of Stock Market --| Crash | Normal |
|------------------------|-------|--------|
| % of N                 | 13.7% | 5.2%   |
| Mean Monthly SP500 Return | 0.17  | -2.78  |
| Median Monthly SP500 Return | 0.9   | -2.46  |
| Stddev of Monthly SP500 Returns | 3.93  | 7.17   |
| Stddev of SP500 Drawdowns | 0.06  | 0.1    |

Date range: 01/01/1959 to 12/01/2020. Positive stock market returns are in green color, and negative returns are in red. Two stock market states (crash and normal) and two economy states (recession and normal) are shown.
Data

ML needs as much data as possible. So, 134 monthly macroeconomic variables based on the FRED database, referred to as the FRED-MD dataset (McCracken & Ng, 2016), are used in this study. A complete definition and listing of the 134 variables are available. This dataset covers the period from 01/01/1959 to 12/1/2020 (or 743 monthly observations).

These 134 variables are selected from eight groups: output and income, labor market, consumption and orders, orders and inventories, money and credit, interest rate and exchange rates, prices, and the stock market. FRED-MD is a modified version of the 132-variable series, sometimes referred to as the ‘Stock-Watson dataset’ (Stock & Watson, 2006). The resulting $743 \times 134$ input vector is classified as a recession and/or stock market crash every month. After computing the recession and drawdowns, one is left with a $731 (743 – 12) \times 136 (134 + 2)$ vector for further analysis.

---

6 All 134 variables: https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/Appendix_Tables_Update.pdf.
The summary data in Table 1 shows that out of 731 months, both the economy and stock market were in a normal state in 528 months (or 72.2% of the time). Only in 38 months (or 5.2% of the time) the economy was in a recession, and the stock market was in a crash period. The stock market yielded an average monthly return of 0.97% when the economy and drawdowns were normal. It generated a $-2.78\%$ monthly return during a recession when drawdowns were severe.

During the recession, one can also notice that drawdowns can be severe: stock markets corrected (peak to trough) by 31% during crash periods. It is also interesting to note that in 100 months, the economy was normal, but the stock market was in a crash state (post-recession, sudden-crash, etc. ex: early 2003, late 2009, late 1987, and 1975).

The historical overview of the U.S. stock market returns and drawdowns in Fig. 1 shows that, as expected, drawdowns increase during crash periods. The 2008 global financial crisis (GFC) experienced the worse drawdown of 51% in March 2009, followed by a 44% drawdown in February 2003 (dot-com crash) and 43% in Dec 1974 (oil crisis and fall of Bretton Woods system).

The Covid-19 crash in March 2020 was neither severe (barely touched the horizontal dotted crash threshold line of 20%) nor long-lasting (the stock market recovered in two months and touched new highs in less than six months). It is also interesting to note that the worse single-month declines occurred recently in October 2008 (a drop of 20.4%) and March 2020 (19.1%). Extreme negative returns are more common than positive returns.

### 4 Methodology and Results

This section explains the ML analysis of recessions and crashes in detail. A modular approach is followed for this analysis using “six steps,” each described in the subsections below.

#### 4.1 Normalization of All Factors (Features)

Standardization of datasets is a common requirement in ML methods. The $743 \times 134$ input vector explained in Sect. (3) contains different scales: interest rates in percentages, industrial production as an index, all employees in millions of value, etc. So, all the values are normalized to have a zero mean and unit variance using the formula below.

$$
\bar{X}_{i,t} = \frac{X_{i,t} - \mu_i}{\sigma_i}, \quad \mu_i = \frac{\sum_{t=1}^{T} X_{i,t}}{T}, \quad \sigma_i = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (X_{i,t} - \mu_i)^2}
$$

(1)

where $\bar{X}_{i,t}$ is a normalized form of feature, $X_{i,t}$, $T$ is the number of observations in the dataset.

As explained next, the full dataset (total rows, $N=743$) is divided into training and testing.
4.2 Training, Validation, and Test Data Selection

The training dataset is used to fit the ML model. The validation dataset is used to tune (penalize) ML model parameters. A ten-fold cross-validation approach is used. Ten data bins are created in each run, nine bins are used for training, and one bin is used for validation. Repeat these runs many times, making ten different random bins (folds) until the model learns to select the key features accurately. The cross-validation approach protects against overfitting, estimates the test error rate by holding out a subset of the training observations from the fitting process, and applies the statistical learning method to those held out observations (James et al., 2013). Data between 01/01/1959 to 07/01/2019 is used for training and validation.

The test dataset, walk-forward out-of-sample, is never part of the training and validation and is used to predict outcomes based on the estimated ML model. The predicted results are often compared with the user’s real-world outcome to see if the ML model performs well in the real world. Data between 08/01/2019 and 12/01/2020 is used for testing.

In Aug 2019, most people did not know about the Covid-19. So, both the people (economy and stock market) and the ML model had no idea of the Covid-19 when the model was estimated. It means that the ML model has to learn about the Covid-19 impact on its own. No data mining (i.e., playing around with the start or end dates) took place to write this paper.

4.3 Drawdown Computations

In this paper, a drawdown refers to the percentage decline of the S&P500 index from the previous peak. A drawdown of more than 20% is classified as a stock market crash. At least 12 months of data is needed to compute drawdowns. The beginning
The drawdown is calculated using Eq. (2). After adding recession and drawdown status as binary features, one is left with a $731 \times 136$ vector for further analysis. A sample drawdown profile of S&P500 from 1990 onwards is shown in Fig. 2. Some drawdowns take years to recover, such as the 2000 crash that took 85 months to recover. In contrast, some take only a few months to reach previous highs, such as the Covid-19 crash, which took five months to recover.

$$DD_t = \frac{\text{Previous Peak Index Value} - \text{Current Index Value}}{\text{Previous Peak Index Value}}$$  \hspace{1cm} (2)

### 4.4 Visualizing Data as a Heatmap

A traditional econometrician will see a fork in methodology from this point onwards. Intuition now takes a back seat as neither humans can visualize a 136 variable heatmap nor traditional statistical methods deal with so many variables. ML algorithms can automatically take lags and process the information to make matters worse. Imagine adding 1/2/3/4/6/12-month lagged variables in traditional estimation.
methods—very quickly, numbers go up from 136 to 816 collinear variables, thus resulting in a near-impossible task to manage the estimation process.

To illustrate the point, eight of the potential 816 features are shown in Fig. 3 as a heat map, which is just one of the many ways ML methods see data. Analyzing data through visualization is a relatively new concept (Friendly, 2008; Helfman & Goldberg, 2014). Studying the stock market through audio analysis could be the next frontier. Researchers have already reported a formal evaluation of visual-auditory display and tested it with nonexperts’ help on their ability to use the new technology to predict stock prices’ future direction (Frylinger, 1990; Nesbitt & Barrass, 2004).

4.5 ML Model Selection and Feature Assessment by Classifier Performance

ML algorithms are numerous: currently, Wikipedia lists 60 ML algorithms,7 and Scikit-learn (Pedregosa et al., 2011) documents 117.8 Soon, they will be in the thousands. It is not easy to know them all and decide what algorithm may work in which scenario. So, ensemble methods,9 using multiple learning algorithms to obtain better predictive performance than could be obtained from any constituent learning algorithms alone, have gained popularity (Opitz & Maclin, 1999; Polikar, 2006; Rokach, 2010). Ensemble methods are more accurate than any single one of the underlying ML methods. The use of Ensemble methods is often supported based on Stein’s paradox. In 1955, Stein discovered that combined estimators are more

---

7 ML algorithm list: https://en.wikipedia.org/wiki/Outline_of_machine_learning#Machine_learning_algorithms.
8 List of ML algorithms at scikit-learn: https://scikit-learn.org/stable/index.html.
9 Ensemble methods: https://en.wikipedia.org/wiki/Ensemble_learning.
accurate on average than any individual estimator they are based on (Efron & Morris, 1977; Stein, 1956).

The ML algorithm selector tool shown in Fig. 4 is used for recession and crash prediction. The following algorithms are recommended: (1) Ensemble learning; (2) linear support vector machine; (3) $k$-nearest neighbors. Also, we chose the most-commonly-used supervised ML classifiers: (4) discriminant analysis; (5) binary decision tree; (6) generalized linear regression model; (7) naive Bayes; (8) decision trees; (9) logistic regression; (10) gradient boosting; (11) AdaBoost; (12) random forest; and (13) XG boost. Prado (2020) describes these methods in detail.

Factor selection is a hotly-debated topic in finance. On the one hand, some researchers prefer meaningful and significant factors. Fama and French (1993) three factors and Fama and French (2015, 2017) five factors that explain average stock returns are some of the notable factors. On the other hand, other researchers believe that factor production in academic research is a “zoo” and out of control (Feng et al., 2020). Harvey and Liu (2019) documented over 400 factors published in top journals. They state that many of those lucky findings were false and resulted from the incentives that lead to factor-mining.

One of the most pervasive mistakes in financial research is to take some data, run it through an ML algorithm, backtest the predictions, and repeat the sequence until a nice-looking backtest shows up. Academic journals are filled with such pseudo-discoveries (Prado, 2018). Unlike the traditional econometric models that explain features in great detail, most ML methods give lower importance to the intuition behind features and higher importance to out-of-sample forecasts on a test dataset. Intuition fails in high dimensions (Domingos, 2012).

Feature selection, ML term for factor selection, is more of a technical exercise than economic analysis. A typical ML model has hundreds of features, and many of them could be collinear. Substitution effects are the ML equivalent of what the statisticians call multi-collinearity. One way to address linear substitution effects is to conduct a principal component analysis (PCA) on the raw features and then select the orthogonal features based on importance. So, most of the time, feature selection in ML analysis is based on non-intuitive PCA factors.

However, by using a technique called maximum relevance and minimum redundancy feature selection method (Hanchuan Peng et al., 2005; Zhao et al., 2019), one could extract the underlying economic factors that best describe the model outcome. Usually, feature importance is estimated from the training and validation dataset.

The 13 ML algorithms are trained and validated on data between 01/01/1959 and 07/01/2019. The criterion was to forecast the following eight binary classifiers (four each for recession and stock market crash): state of the economy (recession or normal) and state of the stock market (crash or normal) in the current month, next month, two months later, and three months later.

In summary, the ML algorithm provides up to a three-month forecast of what may happen in the economy and the stock market. The accuracy of forecasts goes down as the models try to forecast beyond the three-month range. Moreover, a three-month ahead alert of a stock market crash is sufficient time to alert market participants.
Next, features that forecast a three-month-ahead recession are explained first, followed by those that forecast stock market crashes. The same explanation can be extended to the two/one/current month forecasts. One often does not know if the economy is in a recession in the current month since the NBER dating committee announced a recession several months after it had started. That is why the current month’s forecast is also included.

4.5.1 ML Features (Factors) That Predict a Three-Month Ahead Recession

To illustrate an example, all the 13 ML methods are estimated based on the training and validation data from 01/01/1959 to 07/01/2019. The ‘linear support vector machine’ (LSVM) model (Hastie et al., 2009, p. 417) emerged as the best model to predict recessions based on a cross-validation accuracy of 97.8%. The
The ‘bagged ensemble’ method (Hastie et al., 2009, p. 605) is the second-best (validation accuracy of 97.2%). Due to space constraints, charts and graphs for only the first best method (i.e., LSVM) are shown. As shown in Table 1, the bad states (i.e., recession, crash, or failures) occur less than 20% of the time. When failure rates are low, false discovery rates become important. If the model predicts a normal state, but the reality turns out to be a bad state, the consequences of a false prediction can be severe.

The false discovery rate, FDR or a type 1 error, gives the proportion of the incorrectly classified per predicted class. On the other hand, the positive predicted value, PPV, gives the proportion of true positives per predicted positives (FDR + PPV = 100%). The actual and predicted rates by the LSVM method are shown in Table 2. Predicting a recession is three times harder than predicting a stock market crash—the best model’s FDR for three-months ahead prediction is 8.6% for recession and a paltry 2.9% for a stock market crash. Out of the 93 recession months in the training and validation dataset, the LSVM model accurately fit 85 of them (or 91.4% PPV). Similarly, the model accurately fits 626 out of 634 normal economy months (or 98.7% PPV).

In general, all the 13 ML models have performed remarkably well (as measured by accuracy) in predicting in-sample recessions compared to in-sample stock market crashes. The out-of-sample forecasts are shown in the following subsection. The top-five features that predict a recession in three months are shown in Fig. 5, based on model accuracy. The algorithms automatically selected weights (i.e., equal, exponential) on past data. Out of 134 features, 107 have almost no predictive power (i.e., the weight of less than 1%).

As shown in Fig. 5, the top five determinants of a recession are average weekly hours in manufacturing, housing starts in the west, S&P volatility index, S&P dividend yield, and the yield spread between the 1-year Treasury fed funds rate. Some

| Feature group (mapped to FRED-MD) | Feature weight (%) |
|----------------------------------|-------------------|
| Labor market                     | 28.8              |
| Housing                          | 25.0              |
| Stock market                     | 20.6              |
| Interest and exchange rates      | 18.1              |
| Consumption and inventories      | 3.5               |
| Output and income                | 1.7               |
| Money and credit                 | 1.3               |
| Prices                           | 0.9               |
| Grand total                      | 100.0             |

The feature importance scores (as determined by the most accurate ML algorithm, LSVM) are grouped by the FRED-MD categories to calculate weights. Training and validation are conducted from 01/01/1959 to 07/01/2019. The current labor market is the most significant predictor of an impending recession.
of these predictors have appeared in the past literature: yield spread (Moneta, 2005), problems in housing and consumer durables (Leamer, 2008), and the VIX (Stock & Watson, 2012).

The feature importance score of all 134 features is then grouped by the eight categories in the FRED-MD dataset. The resulting feature group weights are summarized in Table 3. The current labor market and housing constitute more than 50% of the weight. It is interesting to notice the contrast between the traditional, intuition-based econometric view of recessions and the high-dimensionality-based ML. Traditional econometricians believe that all the recessions since the mid-1980s have had financial origins (Ng & Wright, 2013). However, the non-intuitive ML algorithms observe the labor markets to foresee a recession. With the help of traditional econometric methods, some researchers found that labor markets and stock prices predicted the 2001 recession (Stock & Watson, 2003). So, one can summarize that according to the ML methods, the labor market and housing are harbingers of a recession in the U.S. ML methods do not have a dogma—so they may look to other areas to foresee the next recession.

A complete heat map of all recession predictors, sized by their importance, is shown in Fig. 6 for completeness. Just the top-five features constitute 53.7% of the total importance.

4.5.2 ML Features (Factors) That Forecast a Three-Month Ahead Stock Market Crash

After repeating the same process as outlined in the previous subsection, the two best ML models for crash prediction are LSVM (same as the best recession predictor model) and ‘decision tree’ (coarse, meaning the maximum number of splits is four). Of all the well-known ML methods, decision trees come closest to meeting

![Fig. 6](image-url)
the requirements for serving as an off-the-shelf procedure for data mining (Hastie et al., 2009, p. 352). The cross-validation accuracy is 95.5% for LSVM and 94.7% for decision tree. Again, charts for only the first-best method (i.e., LSVM) are shown due to space constraints.

ML algorithms predict a crash is relatively easier than predicting a recession—the best model’s FDR for three-months ahead prediction is 8.6% for recession and a paltry 2.9% for a stock market crash. Out of the 137 crash months in the

| Feature group (mapped to FRED-MD) | Feature weight (%) |
|-----------------------------------|-------------------|
| Stock market                      | 30.3              |
| Labor market                      | 30.0              |
| Interest and exchange rates       | 19.0              |
| Output and income                 | 8.3               |
| Consumption and inventories       | 5.7               |
| Prices                            | 4.3               |
| Money and credit                  | 1.7               |
| Housing                           | 0.7               |
| Grand total                       | 100.0             |

All feature importance scores (as determined by the most accurate ML algorithm, linear SVM, that predicted a recession three months down the road) are grouped by the FRED-MD categories to calculate weights. Training and validation are conducted from 01/01/1959 to 07/01/2019. The current labor market is the most significant predictor of an impending recession.
Application of Supervised Machine Learning Techniques to…

training and validation dataset, the LSVM model accurately fit 121 of them (or 88.3% PPV). Similarly, the model accurately fits 562 out of 578 normal economy months (or 97.1% PPV).

The top-five features that determine the stock market crash three months ahead are shown in Fig. 7. Of 134 factors, 101 have almost no effect (weight of less than 1%). The current stock market levels and volatilities are the biggest predictors of a future U.S. stock market crash—both can be reasonably guessed to be significant factors of a stock market crash. Traditional jump literature (Eraker et al., 2003; Santa-Clara & Yan, 2010) may also agree with these findings.

The feature importance score of all 134 features, grouped by the eight FRED-MD dataset categories, is summarized in Table 4. Financial features (stock market, interest rates, exchange rates, money, and credit) contribute to more than half (51.0%) of the weights. One can summarize that financial factors have a bigger role in stock market crashes than economic factors. The labor market appears as a top-two predictor of both recessions and crashes. A complete heat map of all crash predictors, sized by their importance, is shown in Fig. 8.

4.6 Out-of-Sample Forecast of the COVID-19 Recession and Stock Market Crash

This section shows what happens when the rubber hits the road. Having fit the model using training and test data from 01/01/1959 to 07/01/2019, it is now time to test how well the ML models predict the states of the economy and stock market from 08/01/2019 to 12/01/2020. As stated previously, in Aug 2019, most people did not know about the Covid-19. So, the system (people, economy, stock market, and the trained ML model) had no idea of the Covid-19 when the test cycle (out of sample forecasts) began.
One can see that by Feb 2020, the economy had entered a recession, according to the NBER classification. The Stock market crashed by 33.9% (from an S&P500 index closing price high of 3386 on Feb 19th to a low of 2237 on Mar 23rd). The U.S. real GDP decreased by 31.4% in the second quarter (BEA & 2nd quarter, 2020), the sharpest economic contraction in 73 years (Reuters, 2020). The NBER announced in Jun 2020 that a recession had started in the U.S. (NBER, 2020). The monthly data from FRED-MD is still in recession, as announced in Jun 2020. The monthly data from FRED-MD is still in recession, as announced in Jun 2020. Thus, the NBER classification. Keep in mind that NBER had not announced it until Jun 2020 (highlighted in yellow color). We do not know if the Covid-19 recession has ended as NBER has not yet announced that it ended. So, we expect that the economy is still in recession, as announced in Jun 2020. The NBER announced in Jun 2020 that a recession had started in the U.S. (NBER, 2020).

In this backdrop, the two-best ML algorithms saw the economy, as depicted in Fig. 9.

Next, out-of-sample ML forecasts are explained in detail. Focus on the first two columns in Fig. 9: the 1st column (Date) shows the current month, and the 2nd column shows the expected state of the economy in the current month (R0_Expected). One can see that by Feb 2020, the economy had entered a recession, according to the NBER classification. Keep in mind that NBER had not announced it until Jun 2020 (highlighted in yellow color). We do not know if the Covid-19 recession has ended as NBER has not yet announced that it ended. So, we expect that the economy is still in recession, as announced in Jun 2020. The monthly data from FRED-MD is only available until 11/1/2020 at the time of writing this paper. The last column can predict three months into the future (i.e., until March 2021).

Fig. 9  Look ahead, out of sample. Recession forecasts of the best-two ML algorithms. Four grouped-columns show recession forecasts for the current month, next month, two- and three months ahead. Each column contains the expected result (which the author manually added for this paper based on hindsight). The two-best algorithms (by Val_Accuracy) were previously decided on the training data. ML models do not look ahead to forecast. They only have data available up to that month. The forecasts start in Aug 2019 and proceed one month at a time until 12/1/2020. The last column can predict three months into the future (i.e., until March 2021).
what we expect as the predicted state of the economy from the ML algorithms. R0_Expected should be “Normal” from 8/1/2019 to 1/1/2020 and “Recession” from 2/1/2020 till the end.

The 3rd and 4th columns labeled R0_Forecast show the first-best and second-best ML algorithm’s forecast of the current month’s state of the economy. The first-best algorithm to forecast the current month is “KNN Weighted.” It has a validation accuracy (Val_Accuracy) of 96.2% on the training dataset with a PPV of 93.6% (i.e., previously identified recession states accurately 93.6% of the time) and missed them 6.4% of the times (i.e., FDR). The area under the curve (AUC), a measure of a binary classifier’s performance, is 98.0% (i.e., a near-perfect score of 100%). Alas, the 3rd column shows that KNN-weighted, the first-best ML algorithm, could not detect a recession on 2/1/2020 as expected. Neither did the “Ensembled Bagged,” the second-best algorithm. However, a month later, in Mar 2020, the first best detected “Recession,” Eureka moment! Though the second-best still did not see a recession. Another month later, in Apr 2020, both the first- and second-best ML algorithms detected “Recession.”

Continuing further, in the 3rd and 4th columns, one can see that by Jun 2020, both ML algorithms detected that recession was over. So, ML detects that the recession was over, even before the NBER announcement that it had started.

The 5th column is an extension of the 2nd column. The 5th column shows the economy’s expected state next month, “R1_Expected”, or one month ahead. Since “R0_Expected” was “Recession” in Feb, “R1_Expected” should be “Recession”
in Jan. Extending the same logic further, “R2_Expected” should be “Recession” in Dec 2019, and “R3_Expected” should be “Recession” in Nov 2019.

Of all 13 models, the earliest forecast of “Recession” was provided by the “R3_Forecast” from the “SVM Linear” method in the 12th column in Dec 2019. So, the ML methods alerted in December 2019 that a recession would occur in three months in the U.S. Almost all models provided a valuable and accurate forecast of an impending recession.

After the models detect that they have entered a recession state, the next turning point is getting out of it and entering a normal state again. Most ML methods suggest that the U.S. recession has already ended before Jan 2021. The NBER has not yet announced if the recession has ended when writing this paper in Jan 2021.

Next, we switch focus from recession forecasts to crash forecasts, as provided in Fig. 10. To follow easily, Fig. 10 is a replica of the Fig. 9 with only one difference: the letter “R” for recession is replaced by “C” for the crash. Unlike the NBER delayed announcement of a recession, a stock market crash is easy to observe. The yellow color horizontal row shows that in March 2020, S&P500 crashed, or the drawdowns exceeded 20%.

On average, a crash state is more common than a recession state. As shown in Table 1, the economy was in a recession state for 103 months and a crash state for 138 months between 01/01/1959 and 12/01/2020. However, the Covid-19 recession is different—the crash state lasted only one month, and the recession state lasted a lot longer.

The results in Fig. 10 show that ML models provided advanced alerts of a crash. The earliest warning came from both of the “C3-Forecast”, three-month crash alerts issued by “SVM linear” and “Tree Coarse” ML methods in Jan 2020. In essence, the ML methods provided a crash alert more than two months in advance. In general, they detected a shorter crash period. Only one of the models, “C0_Forecast” of the “SVM Linear” method, thought a second crash was coming in Jul/Aug 2020. It was a false alert in hindsight, and hopefully, the model recognized it. ML models learn from their mistakes and get better as they go.

5 Conclusions

Machine learning is a revolutionary technology. One cannot ignore the possibilities and disruptions it can cause. Many economists frown upon ML methods due to ML’s disdain for intuition and focus on high-dimensional predictability. Simultaneously, the financial industry has openly embraced the ML practice; it is one of the hottest career areas in finance. This paper tried to demonstrate that ML can be equally useful in economics and finance with a practical use-case example of recession and stock market crash predictability.

ML algorithms detected the Covid-19 recession six months before the official NBER announcement after learning from strictly past monthly data. Also, they detected a stock market crash two months in advance and alerted us that crash would not last very long. Though the underlying ML factors can change rapidly, the labor
market and housing are harbingers of a future U.S. recession. The labor market appears as a top-two feature in predicting both recessions and crashes.

Currently, ML programming is very computer-science-like. There is a fair amount of objected-oriented programming in this area. Hopefully, tools and technologies will evolve to make it user-friendly and easier to use. The six-step methodology explained in this paper can be applied to many more areas in both economics and finance.

Funding This study is not funded by any entity.

Declarations

Conflict of interest Authors do not have any conflicts of interest associated with this article.

Ethical Approval This article does not contain any studies with human participants or animals.

References

ABC News, A. N. (2020). Cataclysmic event warning. ABC News. https://abcnews.go.com/Politics/intelligence-report-warned-coronavirus-crisis-early-november-sources/story?id=70031273

Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8(2), 937–953. https://doi.org/10.1007/s13198-016-0551-y

Ahrens, R. (2002). Predicting recessions with interest rate spreads: A multicountry regime-switching analysis. Journal of International Money and Finance, 21(4), 519–537. https://doi.org/10.1016/S0261-5606(02)00006-2

Amihud, Y., Mendelson, H., & Wood, R. A. (1990). Liquidity and the 1987 stock market crash. The Journal of Portfolio Management, 16(3), 65–69. https://doi.org/10.3905/jpm.1990.409268

Athey, S. (2018). The Impact of Machine Learning on Economics (No. c14009). National Bureau of Economic Research. https://www.nber.org/books-and-chapters/economics-artificial-intelligence-agenda/impact-machine-learning-economics

Bai, M., Xu, L., Yu, C.-F., & Zurbruegg, R. (2020). Superstition and stock price crash risk. Pacific-Basin Finance Journal, 60, 101287. https://doi.org/10.1016/j.pacfin.2020.101287

BEA, 2nd quarter. (2020). Gross Domestic Product (Third Estimate), Corporate Profits (Revised), and GDP by Industry, Second Quarter 2020|U.S. Bureau of Economic Analysis (BEA). https://www.bea.gov/news/2020/gross-domestic-product-third-estimate-corporate-profits-revised-and-gdp-industry-annual

Bernanke, B. S. (2012). The great moderation. In Book chapters. Hoover Institution, Stanford University. https://ideas.repec.org/h/hoo/bookch/4-6.html

Blanchard, O., & Simon, J. (2001). The long and large decline in US output volatility. Brookings Papers on Economic Activity, 2001(1), 135–174. https://doi.org/10.1353/eca.2001.0013

Borowski, K. (2015). Moon phases and rates of return of wig index on the warsaw stock exchange (SSRN Scholarly Paper ID 2637462). Social Science Research Network. https://papers.ssrn.com/abstract=2637462

Chollet, F., & Allaire, J. J. (2018). Deep learning with R. Manning Publications Co.

Cowles, A. (1933). Can stock market forecasters forecast? Econometrica, 1(3), 309–324. https://doi.org/10.2307/1907042

Croushore, D., & Stark, T. (2001). A real-time data set for macroeconomists. Journal of Econometrics, 105(1), 111–130. https://doi.org/10.1016/S0304-4076(01)00072-0

de Prado, M. (2018). Advances in financial machine learning. Wiley.
Application of Supervised Machine Learning Techniques to...

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking and Finance*, 34(11), 2767–2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

Leamer, E. E. (2008). *Macroeconomic patterns and stories*. Springer.

Leippold, M., Wang, Q., & Zhou, W. (2021). Machine learning in the Chinese stock market. *Journal of Financial Economics*. https://doi.org/10.1016/j.jfineco.2021.08.017

Lowenstein, R. (2004). *Origins of the crash: The great bubble and its undoing*. Penguin.

McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business and Economic Statistics*, 34(4), 574–589. https://doi.org/10.1080/07350015.2015.1086655

McKinsey & Company. (2020). *The state of AI in 2020*. https://www.mckinsey.com/Business-Functions/McKinsey-Analytics/Our-Insights/Global-survey-The-state-of-AI-in-2020

Moneta, F. (2005). Does the yield spread predict recessions in the Euro area?*. *International Finance*, 8(2), 263–301. https://doi.org/10.1111/j.1468-2362.2005.00159.x

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106. https://doi.org/10.1257/jep.31.2.87

NBER. (2020). *Business Cycle Dating Committee Announcement June 8, 2020*. NBER. https://www.nber.org/news/business-cycle-dating-committee-announcement-june-8-2020

Ng, S., & Wright, J. H. (2013). Facts and challenges from the great recession for forecasting and macroeconomic modeling. *Journal of Economic Literature*, 51(4), 1120–1154. https://doi.org/10.1257/jel.51.4.1120

Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11, 169–198. https://doi.org/10.1613/jair.614

Palmer, R. (2008). Angry Queen asks: Why didn’t anyone see the credit crunch coming? Express.Co.Uk. https://www.express.co.uk/news/uk/69678/Angry-Queen-asks-Why-didn-t-anyone-see-the-credit-crunch-coming

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12, 2825–2830.

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *AUDITING: A Journal of Practice and Theory*, 30(2), 19–50. https://doi.org/10.2308/ajpt-50009

Philippon, T. (2016). *The fintech opportunity* (No. w22476). National Bureau of Economic Research. https://doi.org/10.3386/w22476

Pierce, A. (2008). The Queen asks why no one saw the credit crunch coming. The Telegraph. https://www.telegraph.co.uk/news/uknews/theroyalfamily/3386353/The-Queen-asks-why-no-one-saw-the-credit-crunch-coming.html

Pinar Saygin, A., Cicekli, I., & Akman, V. (2000). Turing test: 50 years later. *Minds and Machines*, 10(4), 463–518. https://doi.org/10.1023/A:1011288000451

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3), 21–45. https://doi.org/10.1109/MCAS.2006.1688199

Prado, M. (2020). *Machine Learning for Asset Managers. Elements in Quantitative Finance*. Cambridge University Press.

Professor Arthur Samuel. (1990). Stanford. https://cs.stanford.edu/memoriam/professor-arthur-samuel

Pyle, D., & San José, C. (2015). An executive’s guide to machine learning | McKinsey. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/an-executives-guide-to-machine-learning

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

Reuters, R. (2020). U.S. economy contracts at 31.4% annualized rate in second quarter. Reuters. https://www.reuters.com/article/us-usa-economy-gdp-idUSKBN26L2CC

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1), 1–39. https://doi.org/10.1007/s10462-009-9124-7
Russel, S., & Norvig, P. (2013). *Artificial intelligence: A modern approach* (3rd ed.). Pearson Education Limited.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development, 3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Santa-Clara, P., & Yan, S. (2010). Crashes, volatility, and the equity premium: Lessons from S&P 500 options. *The Review of Economics and Statistics, 92*(2), 435–451. https://doi.org/10.1162/rest.2010.11549

Shapiro, M. D., & Watson, M. W. (1988). Sources of business cycle fluctuations. *NBER Macroeconomics Annual, 3*, 111–148. https://doi.org/10.1086/654078

Shefrin, H. (2002). *Beyond greed and fear: Understanding behavioral finance and the psychology of investing*. Oxford University Press.

Shiller, R. J. (2015). *Irrational exuberance: Revised and expanded* (3rd ed.). Princeton University Press.

Sornette, D. (2017). *Why stock markets crash: Critical events in complex financial systems*. Princeton University Press.

Stein, C. (1956). *Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution*. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. https://projecteuclid.org/euclid.bsmsp/1200510165

Stock, J. H., & Watson, M. W. (1992). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience (No. w4014). *National Bureau of Economic Research*. https://doi.org/10.3386/w4014

Stock, J. H., & Watson, M. W. (2003). *How did leading indicator forecasts perform during the 2001 recession?* (SSRN Scholarly Paper ID 2184936). Social Science Research Network. https://papers.ssrn.com/abstract=2184936

Sorlsson, B., & Watson, M. W. (2006). Chapter 10 forecasting with many predictors. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 1, pp. 515–554). Elsevier. https://doi.org/10.1016/S1574-0706(05)01010-4

Stock, J. H., & Watson, M. W. (2012). Disentangling the channels of the 2007–2009 recession (N.o w18094). *National Bureau of Economic Research*. https://doi.org/10.3386/w18094

Timeline of the COVID-19 pandemic in the U.S. (2021). In Wikipedia. https://en.wikipedia.org/wiki/Timeline_of_the_COVID-19_pandemic_in_the_United_States&oldid=1000968383

Trump, D. (2020). Proclamation on suspension of entry as immigrants and nonimmigrants of persons who pose a risk of transmitting 2019 novel coronavirus. The White House. https://www.whitehouse.gov/presidential-actions/proclamation-suspension-entry-immigrants-nonimmigrants-persons-pose-risk-transmitting-2019-novel-coronavirus/

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

White, E. N. (1990). The stock market boom and crash of 1929 revisited. *Journal of Economic Perspectives, 4*(2), 67–83. https://doi.org/10.1257/jep.4.2.67

Zhang, G. P. (2004). *Business forecasting with artificial neural networks: An overview* [Chapter]. Neural Networks in Business Forecasting; IGI Global. https://doi.org/10.4018/978-1-59140-176-6.ch001

Zhao, Z., Anand, R., & Wang, M. (2019). Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform. ArXiv:1908.05376 [Cs, Stat]. http://arxiv.org/abs/1908.05376

Zhong, X., & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation, 5*(1), 24. https://doi.org/10.1186/s40854-019-0138-0
Publisher's Note  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.