Expert Systems With Applications 199 (2022) 116970

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Machine learning models predicting returns: Why most popular performance metrics are misleading and proposal for an efficient metric

Jean Dessain

IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France

ARTICLE INFO

JEL classification:
C45
C53
G11
G17
N2

Keywords:
Stock return predictability
Machine-learning
Deep learning
Time series forecasting
Performance evaluation criteria
Investment efficiency

ABSTRACT

Numerous machine learning models have been developed to achieve the 'real-life' financial objective of optimising the risk/return profile of investment strategies. In the current article: (a) we present and classify the most popular performance metrics used in 190 articles analysed. We noticed that, in most articles, no attention is devoted to the criteria used to compare the algorithms. (b) We evaluate the ability of the metrics used in the literature to assess the efficiency of algorithms to improve investments results. We demonstrate that many of the most popular metrics, like mean squared error (MSE) or root mean squared error (RMSE), are inappropriate for this purpose while others, like accuracy or F1, are just weak. We explain why risk-adjusted return-based metrics are best-in-class, although they suffer from statistical limitations and do not allow easy comparison of algorithms across assets or over time. (c) We propose a new discriminant metric that measures the efficiency of AI models to optimize the risk-adjusted return, which is statistically more robust, and which can test the effectiveness and the stability of models over time and across assets.

1. Introduction

The finance industry has systematically looked for ways to predict future asset returns, and more generally to predict financial time-series data. The main objective of market practitioners (traders, asset managers, professional or retail investors, risk managers, …) is probably less to predict the effective return of an asset than to predict its sign, either for very short periods of time or for longer horizons. But the task is undoubtedly difficult as markets are volatile and noisy environments, with short-term and long-term fluctuations and huge shifts in volatilities.

1.1. Situation

Artificial intelligence (AI) and its sub-fields of machine learning (ML), deep-learning (DL) and reinforcement learning (RL) have proven to be an attractive framework to perform such tasks. The number of academic studies published on this topic has grown at an exponential rate and a comprehensive review of the literature becomes more and more challenging (Bustos & Pomares-Quimbaya, 2020; Huang et al., 2020; Meng & Khushi, 2019; Ozbayoglu et al., 2020; Sezer et al., 2020), if feasible at all.

Using academic databases[1], literature review of articles and Google Scholar, we searched for articles published between 2010 and June 2021 that present AI-based techniques to predict or classify asset returns or for proposing investment decisions in financial assets. We gathered the articles we found in Google Scholar and from the literature review of previous papers summarizing the state of the science (Bustos & Pomares-Quimbaya, 2020; Huang et al., 2020; Meng & Khushi, 2019; Ozbayoglu et al., 2020; Sezer et al., 2020). To have a relatively homogenous research field, we limited our analysis to papers focusing on stock markets or stock indexes. We excluded papers focusing primarily on other assets like currencies and cryptocurrencies, bonds, commodities, ETFs. We also excluded the articles looking only to predict stock volatility rather than stock price or stock return.

We inspected each article and excluded papers deemed irrelevant or those whose focus was not primarily on stock market prediction and trading. We retained 190 papers and we analysed the performance metrics used for comparing and selecting the best algorithms. We do not pretend to a complete exhaustivity given the current inflation of publications. Nevertheless, with these 190 papers analysed, we gather a fair overview of the recent literature so far.

E-mail address: j.dessain@ieseg.fr.

[1] Science Direct, IEEE, Springler-Link Journal, JSTOR, SSRN, ResearchGate, arXiv.

https://doi.org/10.1016/j.eswa.2022.116970
Received 24 September 2021; Received in revised form 9 January 2022; Accepted 21 March 2022
Available online 24 March 2022

Reviewed papers focus (i) on the independent variables and/or (ii) on the structure of the algorithms used for such prediction or classification of expected returns. The goal of these papers is either to improve the explanatory variables or to have better algorithms to capture the information and patterns hidden in the data, for better investment results. But the criteria used to assess these algorithms is not properly and systematically examined and sometimes not examined at all. Our focus will be on the performance metrics used to assess the efficiency of the algorithms.

1.2. Forecasting process

Our aim is to investigate the pertinence of the performance criteria used to evaluate AI models and to assess their efficiency for comparing these models with the pursued objective to improve the risk/return profile of investment strategies.

The forecasting process, whose detailed organisation may vary, embeds 5 main usual steps:

I. The data gathering;
II. The data preparation;
III. The learning algorithm: that can be tree-based decision, SVM, deep learning or reinforcement learning, and either predict the expected future asset returns, classify them or directly derive an investment decision;
IV. The investment strategy[2]: that converts the predictions/classifications into actions (buy – hold – sell) and leads to positions in assets;
V. The comparison: of the performance of the various algorithms used by the investment strategy.

Out of 190 articles reviewed, 50 compare the performance of the forecasting process with the results of step V. These articles are identified in Appendix 2. with a ‘V’ sign in the right column. The other 140 articles skip the steps IV and V of the forecasting process. They compare the results of the algorithms after step III. Without any translation into an investment decision. These articles are identified in Appendix 2. with a “III” in the right column.

1.3. Contribution

In summary, the contributions of this paper are:

1. We present a very large overview (probably the most comprehensive to date) of papers analysing algorithms of machine learning predicting stock returns between 2010 and June 2021. We classify them by the performance metric used to assess the efficiency of the algorithms.
2. We evaluate the ability of the most standard performance metrics to perform the assessment and we demonstrate that the most popular metrics are at best weak metrics, and even sometimes misleading for the assessment of the reviewed algorithms;
3. We propose a new performance metric that makes the comparison among algorithms predicting returns more accurate, more robust and effective across assets and over time.

From a theoretical point of view, the paper proposes a new angle to analyse popular metrics used for assessing prediction of asset returns. The paper does not address the “efficient market hypothesis”, nor does it test it. Nevertheless, with a new performance metric that compares the efficiency of an algorithm against the Buy & Hold strategy and checks its consistency over time, we provide a useful tool to check whether an algorithm can consistently over time overperform the market return and thereby test the efficient market hypothesis.

From a practical point of view, the paper offers a new metric that is statistically more robust than standard metrics seen in the reviewed articles and that can easily be applied to compare algorithms considered by real-world investors.

The rest of the paper is structured as follows: Section 2 lists and briefly describes the most common performance metrics used in recent articles. Section 3 analyses the capacity of the main metrics to adequately evaluate the performance of the algorithms to maximize return and/or optimize the risk/return. Section 4 proposes a simple and efficient metric to compare the algorithms over time and across assets. Section 5 concludes.

2. Common performance metrics

We reviewed 190 articles presenting either several ML and DL algorithms aiming at predicting future asset returns or RL algorithms proposing investment strategies. The performance metrics found in the analysed articles are very diverse.

2.1. Classification of the observed metrics

Based on the parameter that the metric measures, we propose to classify them as follows:

- Error-based metrics: estimate the performance of an algorithm in measuring the error in prediction between the effective return y_t computed ex-post and the value $\widehat{y}_t$ predicted by the algorithm. These metrics include mean squared error (MSE), mean absolute error (MAE) and evolutions thereof. Botchkarev (2018) presents a typology and analysis of the properties of error-based metrics in machine learning regressions.
- Accuracy-based metrics: measure the accuracy of the class assigned by the algorithm to the predicted return compared to the class of the effective return computed ex-post. The classification can be binary with two classes (positive expected return vs negative expected return, or investment vs no investment) or more complex. Hossin & Sulaiman (2015) propose a review on evaluation metrics for data classification. These metrics are based on confusion matrices, correlation coefficient, …and include R, R^2, accuracy, F1, precision or recall, Matthews correlation coefficient (MCC), etc.
- Investment-based metrics: measure the results derived from an investment strategy proposed by the algorithm with buy-hold-sell signals. These metrics can be subdivided into:
  - o Result-based metrics: measure either the monetary results, the realized return or the risk supported to generate the return (volatility, maximum drawdown, etc.) but do not adjust one by the other.
  - o Risk-adjusted return-based metrics: hereafter also referred to as risk/return-based metrics consider simultaneously the return and the risk of the investment strategy and measure how efficient the algorithm is to generate a return under the constraint of risk and to optimize the risk/return profile. Metrics primarily differ by the way they assess the risk. This class of metrics includes Sharpe, Sortino or Calmar ratios, etc.
- Other “informative metrics”: consider side elements such as the number of trades, the number of days a position is held, the costs of the investment strategy, the CPU/GPU time, etc. More than pure metrics, these elements are rather additional information that complements the previous metric classes. We do not describe these metrics any further.

Appendix 1 provides a review of the performance metrics used in the 190 articles, classified as suggested above.

[2] The RL algorithms tend to merge steps III. and IV. into one single step that directly provides an investment decision (buy – hold – sell).

2.2. Use of performance metrics – summary

Nearly three articles out of four do not use result-based or risk/return-based metrics and rely on error-based and/or accuracy-based performance indicators: 26.3% rely exclusively on error-based measures and 26.8% apply exclusively accuracy-based metrics. 21.6% of the articles apply accuracy-based and error-based metrics.

Error-based and accuracy-based metrics are by far the most popular metrics to assess the ability of algorithms to predict returns and to improve the risk-adjusted return of investment strategies (Table 1). It is critical to test whether the most popular metrics provide a reliable indication of the effective performance of the algorithms. This assessment is presented in section 3.

The detailed view with all the metrics applied per article is provided in Appendix 2. In total, the 190 articles apply 510 performance metrics or 2.68 metrics on average per article.

3. Efficiency of the performance metrics to compare algorithms

In the articles we analysed, a great focus is dedicated to the structure of the AI algorithms predicting the asset returns. Another key focus of most articles is to select the best independent variables to improve the quality of the predictions. Little attention is dedicated to how to measure and compare the performance of algorithms.

3.1. Introduction

Some authors do not explain why they use some specific metric, or how they compute it. Many authors (Bao, et al. 2017; Börjesson & Singull, 2020; Chong et al., 2017a; Das et al., 2018; Dingli & Fournier, 2017b, 2017a; Hiransha et al., 2018; Ji et al., 2021; Kao et al., 2013; Kraus & Feuerriegel, 2017; Ma et al., 2021; Nabipour et al., 2020; Ndikum, 2020; Nikou et al., 2019; Pang et al., 2018; Porshnev et al., 2013) explain how they compute the performance metrics but do not explain why they chose these specific metrics.

Some (de Araújo et al., 2012; Borovkova & Tsiamas, 2019; Nti et al., 2020; de Araújo et al., 2018) of those who justify the use of one or several metrics refer to the argument of "most commonly used". Several authors (Ballings et al., 2015; Ding et al., 2015; Henrique et al., 2018; Mallikarjuna & Rao, 2019) explicitly refer to previous articles.

Eventually, some authors (Aguirre et al., 2020; Carta et al., 2021; Chen et al., 2017a; Chen et al., 2017b; Fischer & Krauss, 2018b; Lim et al., 2019; Lv et al., 2019) explicitly justify their choice for specific performance metrics.

Thakkar & Chaudhari (2021) briefly analyse the methods to compare the performance of algorithms. The analysis is principle-based, and the authors affirm that "the performance can be evaluated using error estimations methods such as RMSE, MSE, MAE, and mean absolute percentage error (MAPE), accuracy and directional accuracy metrics, precision, recall, and F-measure". Thakkar & Chaudhari do not attempt to support this claim with any analysis of any kind.

The objective of the AI algorithms that we analyse, and which is shared by the professional investors, is to optimize the expected return of investments under the constraint of the risks generated by the investment. Our analysis will therefore focus on the ability of metrics to provide a good proxy for the ability of an algorithm to achieve the objective of improving the risk-adjusted return.

Table 1
Types and number of classes per article.

	Accuracy-based only	Error-based only	No result- or risk/return-based	Total
Articles	50	51	142	190
% of all articles	26.3%	26.8%	74.7%	100.0%

3.2. Deductive reasoning

Error-based metrics are among the most popular ones with 187 occurrences[3] in the 190 reviewed articles. Error-based metrics are used in any domain as soon as regressions are involved, but for the specific task considered, error-based metrics suffer from two severe weaknesses:

- while they can easily be applied to regression algorithms, they are less applicable[4] with classification algorithms and are inapplicable with reinforcement learning, making comparison between several types of algorithms not possible. Authors that do apply several classes of algorithms, like Mehtab and Sen (2020), use specific performance metrics for each class or do apply a common metric that is not error-based.
- Error-based metrics will equally consider all errors and will not differentiate an error that triggers a bad decision (a mis-investment resulting in a negative return or a missed opportunity with no investment when the asset has led to a positive return) from an error which has no adverse consequence, leading to a positive return or to a non-investment that avoided a negative return.

Demonstration

Let's assume that, to predict the daily returns of an asset i (r_i) over a period of time, we compare two algorithms A and B. The investment strategy we derive is that if the daily return predicted by algorithm A ($\hat{r}_{Ai}$) or respectively by algorithm B ($\hat{r}_{Bi}$), is positive, we invest in asset i for a period of one day or we hold the position if we already had a position in asset i. If the predicted return, respectively $\hat{r}_{Ai}$ or $\hat{r}_{Bi,}$ is not positive, we do not invest or we close the open position if we had one. We do not factor in transactions fees for our simplified example.

Let's assume that algorithm A predicts a null return every day[5] ($\hat{r}_{Ai} = 0$) while algorithm B predicts a return equal to the double of the effective return[6], i.e., $\hat{r}_{Bi} = 2 * r_i$.

If we compare these two algorithms with an error-based metric, they will be perfectly identical, as the absolute values of the errors are equal: $(r_i - \hat{r}_{Bi})^2 = (r_i - \hat{r}_{Ai})^2$ and $|r_i - \hat{r}_{Bi}| = |r_i - \hat{r}_{Ai}|$. Therefore, we have that $MSE_A = MSE_B$ and $MAE_A = MAE_B$. Similarly, we can prove that $RMSE_A = RMSE_B$ and $NMSE_A = NMSE_B$, $MAPE_A = MAPE_B$, etc. Error-based metrics will be unable to discriminate between the algorithms A and B. But the return generated by the two algorithms will differ significantly:

- Algorithm A will trigger no investment and lead to a return equal to 0[7] (or to the risk-free-rate) and no volatility.
- Algorithm B will generate the perfect investment strategy, with only positive daily returns and no missed opportunity, like a perfect theoretical back-trading.

This simplified example[8] illustrates the fact that all errors are not equal; error-based algorithms do miss this critical element. Error-based metrics could lead to severe misevaluation of the performance of

[3] See Appendix 3. for the overview of the metrics and A.3.2 for the error-based metrics.

[4] Classification algorithms also rely on error or loss functions, like the binary cross entropy, the Hinge loss, the multiclass cross entropy, the Kullback Leibler divergence. But we didn't find evidence that these loss functions have been used to compare algorithms' performance.

[5] This kind of issue is relatively common with deep networks facing vanishing gradient problems.

[6] This situation is more theoretical and is for exemplative purpose only.

[7] Or to a risk-free rate if cash deposited is remunerated.

[8] A different example is provided in Appendix 4, which proves the same inefficiency of error-based metrics.

algorithms. Allowing short selling would not change the conclusion of this deductive reasoning.

Accuracy-based metrics are the most popular with 200 occurrences. They do not suffer from the same lack of precision as they focus on a different criterion: the right or wrong classification of the returns or the right or wrong investment decision. But accuracy-based metrics might miss the magnitude of the relative gain from a good decision versus the magnitude of a loss from a bad decision. Accuracy-based metrics are incapable of capturing whether a misprediction leads to a severe financial consequence or a benign one.

In section 3.3, we empirically test the deductive reasoning, and we test the effectiveness of accuracy-based metrics.

3.3. Empirical analysis

We benchmark each metric with the most applied risk/return performance indicators, Sharpe and Sortino with series of returns generated by AI algorithms.

3.3.1. Methodology

We apply several AI regression algorithms: (i) multi-layer perceptron (MLP), (ii) Long Short-Term Memory neural networks (LSTM), (iii) residual neural networks (ResNet), (iv) Support Vector Machine (SVM) and (v) a decision tree-based algorithm "eXtreme Gradient Boosting" (XGB) to 28 stocks[9] of the DJIU. We use different hyper-parameters[10] with each algorithm to generate 980 series of daily returns. We use 20 years history of daily prices: 15 years are used to train our algorithms and 5 years (1260 days) for testing[11] as out-of-sample data. The independent variables consist of the Open-Close-High-Low prices and traded volumes of the previous days, 14 technical indicators[12] and the close prices of the other stocks of the same index. The algorithm receives end-of-day information and predicts the return for a purchase at the opening of the next day and a sale at the opening of the day after. In total we have 35 different algorithms applied to 28 stocks, and 980 series.

It is important to note that the quality of the algorithms[13] is not relevant for the purpose of this analysis. The importance is to obtain enough different series of daily returns for which we compute the various performance metrics to assess their efficiency.

We compute the MSE, RMSE, MAE and MAPE of the regressions. We benchmark each of the 980 series with the "back-trading" of a perfectly informed agent that invests when the return is positive and doesn't invest when the return is negative or zero. We compute R, R^2, accuracy, F1, precision & recall and Matthew's correlation coefficient (MCC).

We apply the following investment strategy: if the predicted return of the next day is positive, we invest for one day, otherwise we take no open position. In each case, the model integrates direct transactions costs[14] of 0.10% per transaction applied to the value of the transaction. From that investment strategy and assuming a risk-free rate at 0.0%, we compute the annual return (RoI), the volatility (Vol), the yearly maximum drawdown (MDD) in percentage of the investment and the Sharpe,[15] Sortino and Calmar ratios.

We compute first the correlation between the error-based and accuracy-based metrics with the RoI, the Sharpe, Sortino and Calmar ratio, and verify the significance of the correlations. We also perform a series of linear regressions: the performance metrics are the explanatory variables, and the dependent variable is respectively the annualized return and the Sharpe ratio.

3.3.2. Correlation of the metrics

With the error-based metrics, we expect a negative correlation with the RoI, Sharpe, Sortino and Calmar ratios: the lower the error, the better the expected result. In italic, the metrics that are positively correlated.

Against expectations for efficient metrics, correlations disclosed in Table 2 are positive, except between MAPE and the risk/return performance metrics, but the correlations are not significantly different from 0 at 5% significance level, as illustrated with the p-values. MAPE is the only metric whose correlation is negative and significantly so.

Efficient accuracy-based metrics should have positive and significant correlations with RoI, Sharpe, Sortino and Calmar ratios and the accuracy-based metrics, that is higher accuracy. In italic, the negative correlations.

R and R^2 are negatively correlated with the annual return and with the risk/return ratios but not significantly different from zero. Accuracy, F1, precision & recall and MCC are positively correlated with the RoI and with Sharpe, Sortino and Calmar ratios. Accuracy and F1 have the highest correlation and MCC the lowest one. Table 3 presents these results.

3.3.3. Linear regressions

We perform a series of regression $y_i = a_i + b_{ij}X_j$ where i is respectively the RoI and Sharpe ratio, and j each of the error-based and accuracy-based metrics. For each regression, we compare the R^2, the part of the variability in the dependent variable explained by the explanatory variable; the sign of b_{ij} and its significance.

Table 4 shows that R^2 of the linear regressions where error-based metrics are explanatory variable are very low, with RoI and with Sharpe as dependent variables. This result confirms the inability of these metrics to provide reliable information on the capacity of an algorithm to predict a return of a stock.

Correlation coefficient R and determination coefficient R^2 are inappropriate performance metrics with no capacity to assess the performance of the tested algorithms.

From our 980 tests, Table 5 shows that accuracy and F1 have the highest explanatory capacity with R^2 respectively above 45% and 38%. This is still low to assess the performance of the algorithms to predict or classify future returns adequately. MCC is lagging far behind precision and recall to identify the best algorithm.

We performed the same analysis with randomly generated series of returns and obtained very similar results. These are not disclosed here and can be obtained from the author.

3.4. Concrete examples for illustrative purpose

To make it very concrete and easy to visualize, we prove the inefficiency of the error-based and accuracy-based metrics to identify the most efficient algorithm, with a possible financial impact if it would have been applied by a market practitioner investing according to the algorithms.

We compare the results of 3 machine learning algorithms (MLP,

[9] Two stocks out of the 30, Dow and Visa, are not considered for this analysis as the size of available historical data for these stocks do not match the minimum length of 20 years.

[10] size and number of hidden layers, number of epochs.

[11] We do not use the traditional split between training set, validation set, and test set as we are just looking for algorithms to produce series of returns.

[12] 5 days moving average; 12-, 26- and 50-days exponential moving average, MACD Moving Average Convergence Divergence, Bollinger band up & down 20 days, CCI Commodity Channel Index 14 days, ATR Average True Range 10 days, ADX Average Directional moving indeX 5 and 14 days, RSI Relative Strength Index 5 and 14 days and momentum 1 day.

[13] Finding efficient algorithms is the topic of a research that is currently performed and that will be shared later.

[14] The investment strategy integrates transaction costs of 0.10% per transaction. A transaction occurs either when the previous position was 0 and that the algorithm triggers a buy order or when the previous position was positive, and that the algorithm predicts a negative return and a sale.

[15] As the reference rate for the period has been set to 0.00%, this is also equal to the information ratio.

Table 2
Correlation matrix with error-based metrics and p-values.

CORREL	RoI	Sharpe	Sortino	Calmar	MSE	RMSE	MAE	MAPE
RoI	100.00%	98.14%	98.41%	94.50%	*9.60%*	*12.27%*	*10.05%*	−27.99%
Sharpe	98.14%	100.00%	99.74%	96.69%	*8.51%*	*10.65%*	*7.99%*	−27.95%
Sortino	98.41%	99.74%	100.00%	97.48%	*8.88%*	*11.19%*	*8.55%*	−26.75%
Calmar	94.50%	96.69%	97.48%	100.00%	*6.77%*	*9.14%*	*5.45%*	−26.56%

p-value	RoI	Sharpe	Sortino	Calmar	MSE	RMSE	MAE	MAPE
RoI		0.00%	0.00%	0.00%	*9.54%*	3.28%	*8.06%*	0.00%
Sharpe	0.00%		0.00%	0.00%	*13.93%*	*6.42%*	*16.55%*	0.00%
Sortino	0.00%	0.00%		0.00%	*12.30%*	*5.17%*	*13.77%*	0.00%
Calmar	0.00%	0.00%	0.00%		*24.03%*	*11.22%*	*34.42%*	0.00%

Table 3
Correlation matrix with accuracy-based metrics and p-values.

CORREL	R	R^2	Accuracy	F1	Precision	Recall	MCC
RoI	*−2.56%*	*−8.81%*	67.41%	61.89%	58.58%	59.21%	32.65%
Sharpe	*−3.09%*	*−8.86%*	68.80%	63.06%	58.19%	60.51%	34.99%
Sortino	*−1.98%*	*−7.67%*	68.19%	61.19%	57.80%	58.61%	35.09%
Calmar	*−1.92%*	*−7.20%*	64.42%	58.49%	53.55%	56.37%	32.05%

p-value	R	R^2	Accuracy	F1	Precision	Recall	MCC
RoI	*65.66%*	*12.60%*	0.00%	0.00%	0.00%	0.00%	0.00%
Sharpe	*59.17%*	*12.39%*	0.00%	0.00%	0.00%	0.00%	0.00%
Sortino	*73.11%*	*18.30%*	0.00%	0.00%	0.00%	0.00%	0.00%
Calmar	*73.91%*	*21.11%*	0.00%	0.00%	0.00%	0.00%	0.00%

Table 4
Regression with error-based metrics.

Regression results	MSE	RMSE	MAE	MAPE
R^2 vs RoI	0.9%	1.5%	1.0%	7.8%
Coeff. vs RoI	42.37	2.28	3.15	−0.03
p-value vs RoI	0.86	0.00	0.00	0.00
R^2 vs Sharpe	0.7%	1.1%	0.6%	7.8%
Coeff. vs Sharpe	5.28	208.09	10.97	13.84
p-val vs Sharpe	0.81	0.01	0.00	0.01

LSTM and ResNet) applied to the 28 stocks DJIU described supra, and present 3 of them for illustrative purpose, respectively Apple (AAPL), Boeing (BA) and JP Morgan (JPM), over the period 2016–2020. Details are provided in Table 6 of section 4.2.2. Sharpe, Sortino and Calmar ratios provide the same results for the most efficient algorithm with the three stocks. Error-based metrics manage to find the best result for only one stock out of three. Even with BA where LSTM1 algorithm provides the highest RoI and the lowest volatility, error-based metrics fail to identify it. Accuracy, F1, precision and MCC do identify the most efficient algorithm with AAPL but fail with BA and JPM.

These poor results from concrete examples illustrate why the general conclusion from sections 3.2 and 3.3 does matter in real life:

- Error-based metrics, R and R^2 do not provide a reliable way of assessing the quality of the data and the efficiency of algorithms whose goal is to improve investment strategies. They provide misleading indications and should not be used for such purpose.
- Accuracy, F1, precision or recall and MCC provide a reasonably acceptable method for benchmarking algorithms but are no way near the efficiency of the risk/return-based metrics and regularly fail to identify the best performing algorithms.
- Sharpe, Sortino or Calmar ratios should be preferred to accuracy-based metrics as they provide more granularity about the quality of the results. They are the "*best-in-class*" of the current literature reviewed, despite some weaknesses discussed in 4.

4. Proposed new risk/return performance metric

Sharpe ratio dominates risk/return-based performance metrics, far ahead of Sortino, Calmar and information ratios. Sharpe and Sortino ratios suffer from two important issues: (i) they both assume a Gaussian distribution of the returns, and (ii) they do not allow the performance of different algorithms to be compared over different assets or over different time periods. The results of Sharpe and Sortino are influenced by the return of the underlying asset. We propose a new performance metric that improves the risk measurement (in 4.1) and which has the ability to compare the efficiency of algorithms over time and across assets (in 4.2).

Table 5
Regression with accuracy-based metrics.

Regression results	R	R^2	Accuracy	F1	Precision	Recall	MCC
R^2 vs RoI	0.1%	0.8%	45.4%	38.3%	34.3%	35.1%	10.7%
Coeff. vs RoI	*−0.10*	*−3.92*	3.94	0.80	2.17	0.44	1.17
p-value vs RoI	0.42	0.01	0.00	0.00	0.00	0.00	0.00
R^2 vs Sharpe	0.1%	0.8%	47.3%	39.8%	33.9%	36.6%	12.2%
Coeff. vs Sharpe	*−0.67*	*−21.81*	22.25	4.49	11.92	2.48	6.95
p-val vs Sharpe	0.33	0.01	0.00	0.00	0.00	0.00	0.00

Table 6
D-ratio analysis of 3 different algorithms with 3 stocks – best algorithm per stock and per metric is in bold.

	AAPL				BA				JPM			
	B&H	MLP	LSTM	ResNet	B&H	MLP	LSTM	ResNet	B&H	MLP	LSTM	ResNet
RoI	**31.9%**	30.2%	28.8%	27.8%	**7.6%**	2.8%	**6.7%**	5.6%	**12.4%**	1.9%	**4.6%**	4.5%
Vol	29.8%	**19.4%**	21.0%	1.3%	47.4%	26.9%	**25.5%**	26.0%	28.3%	**19.5%**	19.8%	20.0%
D-ratio	1.00	**1.71**	1.44	1.37	1.00	0.90	**2.27**	1.90	**1.00**	0.24	**0.64**	0.63
D 1st	*1.00*	***2.32***	*0.80*	*0.55*	*1.00*	*0.99*	***1.20***	*1.08*	***1.00***	***0.34***	*0.50*	*0.33*
D 2nd	*1.00*	*1.69*	***1.73***	*1.74*	*1.00*	*−0.52*	***−0.06***	*−0.07*	*1.00*	*−0.23*	*0.74*	***1.31***
D-Return	**1.00**	**0.95**	0.90	0.87	**1.00**	0.36	**0.87**	0.74	**1.00**	0.15	**0.37**	0.36
D-VaR	1.00	**1.81**	1.59	1.57	1.00	2.50	**2.59**	2.57	1.00	1.58	**1.74**	1.73
Sharpe	1.069	**1.561**	1.372	1.309	0.161	0.103	**0.262**	0.218	**0.440**	0.097	**0.232**	0.227
Sortino	1.699	**2.576**	2.217	2.108	0.236	0.155	**0.400**	0.332	**0.692**	0.147	**0.363**	0.356
Calmar	0.848	**1.079**	0.968	0.931	0.102	0.066	**0.193**	0.161	**0.298**	0.070	**0.200**	0.197
MSE		0.000	**0.000**	0.000		0.001	0.001	**0.001**		0.000	**0.000**	0.000
RMSE		0.019	**0.019**	0.019		0.031	0.031	**0.031**		0.018	**0.018**	0.018
MAE		0.013	**0.013**	0.013		0.017	0.017	**0.017**		0.012	**0.012**	0.012
MAPE		1.590	1.570	**1.570**		1.769	1.758	**1.757**		1.504	**1.488**	1.492
ACC		**0.553**	0.549	0.548		**0.515**	0.510	0.509		**0.511**	0.509	0.507
F1		**0.641**	0.641	0.640		**0.573**	0.570	0.569		**0.576**	0.576	0.576
Precision		**0.589**	0.584	0.583		**0.537**	0.533	0.532		**0.528**	0.525	0.524
Recall		0.704	0.710	**0.710**		**0.615**	0.612	0.612		0.635	0.638	**0.640**
MCC		**0.058**	0.046	0.042		**0.014**	0.004	0.000		**0.004**	−0.002	−0.006

4.1. *Improving risk measurement*

Sharpe ratio quantifies risk[16] using the standard deviation of excess returns, and Sortino by using the standard deviation of the negative excess returns. They assume that returns are normally distributed, with no skewness and a kurtosis around 3. If a portfolio's return does not follow a Gaussian distribution, then the classical return volatility is no longer an effective measure of risk, and these ratios could underestimate the risk. In addition, Marquering and Verbeek (2004) add that the ratios do not adequately measure the risk-adjusted returns in presence of time-varying volatility.

We tested the normality hypothesis on the data set used in 3.3 with the Shapiro-Wilk test for each series. No series of returns passes the Shapiro-Wilk test at 0.05. Even if we apply the "*back-of-the-envelope*" normality test of a skewness between −0.5 and +0.5 and a kurtosis between 1 and 5, no series among the 980 generated by the algorithms passes the test, and only 25% of the Buy & Hold series could be considered "reasonably" as Gaussian according to this "simplified test".

Value-at-risk (VaR) is another popular measure of the financial risk[17] that offers a way to address skewness and kurtosis of the asset returns distribution with Cornish Fisher expansion (CF expansion). CF expansion accounts for the 4 moments of the distribution: the mean, the volatility, the skewness and the kurtosis. It offers an easily implementable parametric form that improves risk measurement. Cornish-Fisher VaR (CF-VaR) is an effective and easy-to-implement approach to dealing with non-Gaussian distributions. Maillard (2012) proposes a User's guide to the Cornish-Fisher expansion where he describes the CF-VaR and its limits. The Cornish-Fisher expansion is effective within a domain of validity (Amédée-Manesme et al., 2019), outside of which it can lead to mis-estimate of quantiles.

We therefore tested[18] our series of daily returns generated by the algorithms: 84,1% of these series are well within the domain of validity, and 15,9% of the series of our sample have a possibly underestimated risk, but to a lesser extent than with Sharpe or Sortino. There is no easy and parametric way to improve the risk measurement besides CF-VaR. We prefer the CF-VaR to the conditional value-at-risk (Co-VaR) for one main reason: the sensitivity to estimation errors. Co-VaR is statistically more coherent than CF-VaR as a continuous and convex function (Rockafellar & Uryasev, 2002). However, the challenge with Co-VaR is that it is "more sensitive than CF-VaR to estimation errors" (Sarykalin et al., 2008). Therefore, the use of Co-VaR values may prove to be misleading.

Even if the solution is not entirely satisfactory, the use of CF-VaR significantly improves the measure of risk, compared to the volatility of returns. In view of a wide adoption and a relative ease of use coupled with an easy interpretability of the metric, we have chosen to keep the CF-VaR as risk metric. Testing the effectiveness of CF-VaR on return distributions with extreme skewness and kurtosis or with multi-modal distributions is outside the scope of this paper. It is however a perspective to be investigated for further improvements.

If we combine the asset return with the CF-VaR, we can easily define a Return-to-VaR ratio "RtV" equal to RoI / CF-VaR. This ratio outperforms Sharpe, Sortino or Calmar ratios as it better captures the effective risk accepted to generate the effective return. RtV does not assume Gaussian distribution.

While being an improvement compared to the ratios currently used, RtV ratio does not capture the true merit of the algorithm. The value of the ratio is significantly impacted by the asset return that the algorithms try to predict. There is no way to check its stability and efficiency through time nor is it possible to verify that the algorithm is equally efficient for various assets. We propose therefore to remedy to these issues in 4.2.

4.2. *"D-ratio" to capture the algorithm's true added value*

4.2.1. *Principle*

Traditional return-based metrics and risk-adjusted return-based metrics like average annual return or Sharpe ratio do not allow easy comparison over time or across asset. They are measuring absolute levels of return or absolute levels of risk-adjusted return. For example, we cannot assess an algorithm achieving a 8% return or a Sharpe ratio of 2.0. It would be considered as outstanding if the Buy & Hold strategy leads to a 6% annual return or a Sharpe ratio of 0.8. It would be considered as relatively poor performer if the Buy & Hold strategy achieves a 10% annual return or a Sharpe ratio of 2.5. If we obtain the

[16] Calmar ratio quantifies the risk with the maximum drawdown, also referred to as expected shortfall (Auer & Schuhmacher, 2013). Auer & Schuhmacher also cites the Sterling, Burke, Pain and Martin ratios that no article we analysed refers to.

[17] VaR is now mandatory measure for assessing the capital adequacy of Financial Institutions. NB. if the returns are normally distributed, the Sharpe ratio is equivalent to a ratio "return / VaR", with VaR being computed at 84.1% confidence interval.

[18] Amédée-Manesme et al. (2019): with s= skewness and k = excess kurtosis (kurtosis −3):$S^2/9 + 4 * (k/8 – s^2/6) * (1 - k/8–5 * s^2/36)$ is negative when the series is within the domain of validity of the CF expansion.

same 8% return and 2.0 Sharpe ratio over two periods, is the algorithm efficient over the two periods, or highly efficient during one period and poorly performing during the other period?

To address this issue, we propose to work with a relative performance metric that we call the "Discriminant ratio" or "D-ratio", and that we will express in formulas as "D". To build this D-ratio, we will use two components: (i) with the "D-return" ratio, we will compare the annual return of the algorithm to the Buy & Hold strategy. (ii) With the "D-VaR" ratio, we will measure the relative ability of the algorithm to reduce the risk compared to the risk of a Buy & Hold investment. The final metric D-ratio is the relative risk-adjusted return performance of the algorithm compared to Buy & Hold, obtained in combining the D-return ratio with the D-VaR ratio.

4.2.1.1. Measure of the relative return of the algorithm with the D-return ratio. We can compare the annual return generated by an algorithm with the annual return of the Buy & Hold strategy: $RoI_{algo} / RoI_{B\&H}$. If the ratio is above 1, the algorithm generates an average annual return that exceeds the return from the Buy & Hold. We propose to call the overperformance of the algorithm compared to Buy & Hold the "D-return ratio", denominated in formulas as "D-return".

An improvement is required to adequately address the situation where the return of the Buy & Hold strategy and the return of the algorithm are of opposite signs. In the case of opposite signs for the Buy& Hold return versus the algorithm return, we propose to improve the computation as follows:

D-return = 1 + (RoI_{algo} - $RoI_{B\&H}$) / ABS($RoI_{B\&H}$) (1).

When D-return is equal to 1, the algorithm and the Buy & Hold strategy deliver the same annual return. If D-return is below 1, the annual return of the algorithm is below the annual return of the Buy & Hold by a factor equal to 1- D-return. If D-return is above 1, the annual return of the algorithm exceeds the annual return of the Buy & Hold by a factor equal to D-return – 1.

4.2.1.2. Measure of the relative risk of the algorithm with the D-VaR ratio. We compare the risk of the investment strategy of the algorithm with the risk of the Buy & Hold strategy using CF-VaR[19]. We therefore divide the CF-VaR of the Buy & Hold ($VaR_{B\&H}$) by the CF-VaR of the algorithm (VaR_{algo}), and we call this ratio the D-VaR ratio, expressed as D-VaR in formulas:

D-VaR = $VaR_{B\&H}$ / VaR_{algo} (2).

When D-VaR is equal to 1, the CF-VaR of the algorithm is equal to the CFVaR of the Buy & Hold strategy. If D-VaR is below 1, the risk of the algorithm measured by its CF-VaR is higher than the risk of the Buy & Hold by a factor equal to 1- D-VaR. If D-VaR is above 1, the risk of the algorithm is lower than the risk of the Buy & Hold by a factor equal to D-VaR – 1.

One attention point: if the returns are not within the domain of validity described in 4.1 and if there is a difference in skewness and kurtosis between the returns from the algorithm and the returns of the Buy & Hold, D-VaR is not invariant anymore with the confidence interval. As described in 4.1, 84.1% of the returns tested are within the domain of validity.

4.2.1.3. Measure of the relative performance of the algorithm. We therefore propose to define the new relative risk-adjusted return ratio, that we call "**Discriminant ratio**" or "**D-ratio**", denominated as "D" in our formulas. This D-ratio solely focuses on the added value of the algorithm compared to Buy & Hold strategy. D-ratio is therefore a relative measure of performance. To achieve this objective, we multiple the D-return ratio by the D-VaR ratio:

D = D-return * D-VaR (3)[20].

When the D-ratio is equal to 1, the algorithm and the Buy & Hold strategy deliver the same risk-adjusted return. If the D-ratio is below 1, the algorithm underperforms the Buy & Hold strategy from a risk-adjusted return perspective. If the D-ratio is above 1, the risk-adjusted return of the algorithm exceeds the risk-adjusted return of the Buy & Hold strategy.

Structured as a relative performance metric, the Discriminant ratio is therefore not impacted by the return and the level of risk of the underlying asset over the considered period:, the ratio indicates whether the algorithm overperforms the Buy & Hold strategy over the analysed period or not, and to what extent.

As the D-ratio is the product of D-return and D-VaR, we can assess the magnitude by which the algorithm overperforms the Buy & Hold strategy, with D-return the part of the overperformance brought by the algorithm that comes from its ability to increase the return, and with D-VaR the part of the value that comes from the risk reduction capacity of the algorithm.

Relative metrics compared to a reference Buy & Hold strategy, the D-ratio, D-return and D-VaR ratios are independent from the performance of the underlying asset and express the sole merit of the algorithm, contrary to what absolute metrics like average annual return, Sharpe, Sortino do.

The merits of our proposed D-ratio can be regrouped in two categories:

A. The D-ratio better captures the risk of the investment strategy as it is not limited by the assumption of a Gaussian distribution of the returns. This improvement comes from the use of the Cornish-Fisher expansion to refine the Value at Risk computation;
B. The D-ratio is highly versatile:
   a. It is valid for all kinds of algorithms: ML, DL and RL, with regressions or classification. The D-ratio does not measure the direct output of the algorithm, it measures the financial efficiency of investments induced from the algorithm's prediction, it can therefore be used with any kind of algorithm;
   b. The D-ratio is time-insensitive: as the D-ratio is a relative metric, the efficiency of the algorithm can be compared over various periods of time. The stability of the algorithm can easily be verified by testing the D-ratio over the complete period versus two or more sub-periods. This merit comes from its relative character, whereas absolute metrics are influenced by the return and the volatility of the asset over the analysed period.

The stability of AI models is a constant point of attention, this feature is therefore key for assessing the effectiveness of AI algorithms and to avoid non reproducible results. *In our numerical example, the D-ratio would be computed on the entire 5 years period and on two sub-periods of 2.5 years.*

   c. The D-ratio can be decomposed into a first sub-ratio D-return dedicated to the efficiency of the algorithm to improve the return, and a second sub-ratio D-VaR that assess the efficiency of the algorithm to reduce the risk.
   d. The D-ratio allows to compare the algorithms applied with a long only strategy or with short-selling strategies. It allows to measure easily and efficiently the impact of transaction costs on the effectiveness of the algorithm to improve the risk/return of the investment strategy.

[19] For a critic of CF-VaR and its limits, we refer to section 4.1.

[20] D-ratio can also be expressed as (1 + (RoI_{algo} - $RoI_{B\&H}$) / ABS($RoI_{B\&H}$)) * $VaR_{B\&H}$ / VaR_{algo}.

e. The D-ratio allows to compare the efficiency of the algorithm with various assets, from the same asset class (here, stocks) or across various asset classes.

The code in Python for the computation is available on GitHub[21].

4.2.2. Example

We analyse the D-ratio with the same sample as in 3.4. and show here the results of the same 3 stocks for illustrative purpose. On top of comparing the algorithms with the D-ratio, we divide the sample into two sub-periods of 2.5 years and we check the stability of the D-ratio over time with D-1st as the D-ratio for the first sub-period and D-2nd as the D-ratio for the second sub-period. Eventually, we compute the D-return and D-VaR to analyse the relative efficiency of each algorithm to improve the return or to reduce the risk.

From the proposed example, the D-ratio demonstrates that no tested algorithm emerges as stable over time and equally efficient with the three stocks (algorithms applied for illustrative purpose[22] are very simple ones). In this example, no algorithm is stable over the two sub-periods and efficient with the three stocks. The D-ratio would allow to investigate why an algorithm performs better with one stock rather than with another one and to look for algorithm optimization. We would have to improve these algorithms to come up with an attractive solution for real-world investors, and the D-ratio helps us to reach this conclusion, whereas no other performance metric provide such conclusion.

The Table 6 also presents the numerical results of the error-based and accuracy-based metrics that are discussed in 3.4.

The D-ratio proves its efficiency to discriminate between algorithms and to measure the value the algorithms compared with the Buy & Hold benchmark strategy.

5. Conclusion and perspectives

In this final section, we conclude on the results from our analysis. We then draw some perspectives for future research.

5.1. Conclusion

Most of the current literature applies error-based and/or accuracy-based metrics that are easy to use but inefficient to assess the performance of analysed algorithms, leading to unreliable conclusions that can produce severely distorted results. Even the best-in-class metrics (Sharpe and Sortino ratios) suffer from the assumption of normally distributed returns, which is not verified in practice. Furthermore, they do not measure the added value of the algorithms compared to Buy & Hold strategy, but they mix the impact of the algorithm and the intrinsic performance of the underlying asset for the out-of-sample period analysed.

We propose an alternative metric, the D-ratio, that measures the relative under/over-performance of the tested algorithm compared to Buy & Hold strategy rather than a risk-adjusted return of the algorithm. Therefore, the D-ratio is not time sensitive and can be used to compare the efficiency of the algorithm over time or across assets. The D-ratio makes it possible to test the reliability, the reproducibility and the stability of the algorithms, a pre-requisite for the effective adoption of any algorithm in 'real-life'.

While it does not pretend to capture the entire risk in all circumstances, the D-ratio is an improvement compared to current best practice metrics, as it better captures market risks than Sharpe or Sortino ratios or VaR. and as the D-ratio is an improved practical and easy to implement performance metric.

5.2. Perspectives

First, the D-ratio has been applied to simple standard algorithms of ML, DL and RL. The purpose was not to test the best algorithms but to verify the ability of the D-ratio to compare algorithms. The D-ratio should be tested with many other algorithms: the algorithms proposed in the 190 articles reviewed could be re-evaluated with the D-ratio to verify their true effectiveness. The ability of these algorithms to cope with various asset classes and to present stable results over time could also be tested with the D-ratio, as robustness and stability are a prerequisite for any real-life application.

This paper is oriented towards algorithms directly predicting asset returns. Similar empirical analysis could be performed for algorithms predicting asset prices rather than asset returns.

By using the CF-VaR, the D-ratio used an improved risk measure compared to Sharpe or Sortino ratio. The domain of efficiency of the Cornish-Fisher expansion is limited and CF-VaR is not the most adequate risk measure for return distributions with extreme skewness and/or kurtosis, or with multi-modal distributions. Therefore, the D-ratio should be tested with return distributions with very high skewness and/or kurtosis, or with multi-modal distributions, and its sensitivity to the confidence interval of the VaR should be further analysed.

While the D-ratio does not test or discuss the "efficient market hypothesis", it could however be a useful tool to test it, as it measures the ability of algorithms to overperform or not Buy & Hold strategy consistently over time.

The D-ratio is suitable for comparing algorithms predicting asset returns or portfolio returns. It does not provide as such a way to optimize portfolio composition. This might be a topic for further research.

The D-ratio should be suitable for any asset class and for any period of time. It would be useful to test the D-ratio on various assets and on different trading patterns (from intra-day to long-term investments).

Eventually, the D-ratio might serve in other domains than the comparison of AI algorithms predicting returns. For example, comparing fund managers performances and their consistency over time is an area where the D-ratio and its effectiveness could be tested.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix 1. Metrics used

In bold, the main metrics per class.

[21] https://github.com/JDE65/D-ratio.

[22] A paper will follow comparing more efficient algorithms, using the quality of information provided by the D-ratio, that Sharpe, Sortino or Calmar are unable to provide.

Table A.1. List of identified accuracy-based metrics and number of occurrences

Abbreviation	Accuracy-based Metric	Occurrences
ACC	**Accuracy**	**74**
ARV	Average relative Variance	2
AUC	Area under the curve	8
AUROC	Area under the RO curve	1
COV	Covariance	1
DPA	Directional predictive accuracy	8
EF	Evaluation function	2
EV	Explained Variation	1
F1	**F1 or F-measure**	**22**
G-Mean	Geometric mean	1
HIT	Hit rate	1
HM	Henriksson and Merton	1
IC	Information coefficient	1
MCC	Matthew's correlation coefficient	3
MME	Mean of misclassification error	0
MI	Mutual information	2
POCID	Prediction of change in direction	4
Prec	**Precision**	**19**
PT	Pesaran Timmerman test	1
R	Correlation coefficient	13
R^2	Determination coefficient	11
Rec	**Recall**	**19**
ROC	Receiver operating characteristic	2
SAR	SAR Score (ACC + AUC - RMSE)	1
Spec	Specificity	1
WR	Win ratio	1
Total	**Accuracy-based**	**200**

Table A.2. List of identified error-based metrics and number of occurrences

Abbreviation	Error-based Metric	Occurrences
BCE	Binary Cross-entropy	1
LogLoss	Logarithm of the loss	1
MAE	**Mean absolute error**	**37**
MAPE	**Mean absolute percentage error**	**44**
MASE	Mean absolute scaled error	1
MSE	**Mean squared error**	**34**
NMAPE	Normalized MAPE	0
NMSE	Normalized MSE	2
NRMSE	Normalized root MSE	1
RMSE	**Root MSE**	**54**
RMSPE	Root mean squared percentage error	2
RMSRE	Root mean squared relative error	1
RSE	Relative squared error	1
SMAPE	Symmetric MAPE	1
TheilU	Theil's inequality coefficient	7
Total	**Error-based**	**187**

Table A.3. List of identified result-based metrics and number of occurrences

Abbreviation	Result-based Metric	Occurrences
Alpha	Jensen's alpha	1
APL	Average profit / average loss	2
ARet	**Annual return**	**35**
CRes	**Cumulative result**	**17**
MDD	Maximum drawdown	11
Vol	Volatility	13
Total	**Result-based**	**79**

Table A.4. List of identified risk/return-based metrics and number of occurrences

Abbreviation	Risk/return-based Metric	Occurrences
CR	Calmar ratio	5
IR	Information ratio	4
MAR	Minimum acceptable return	2
SorR	Sortino ratio	4
SR	**Sharpe ratio**	**21**
Total	**Risk/return-based**	**36**

Table A.5. Summary of the relative use of the performance metrics

Class of Metric	Occurrences
Error-based	187
Accuracy-based	200
Result-based	79
Risk/return-based	36
Other informative metrics	*8*
Total metrics	**510**
Average per article	**2.68**

Appendix 2. Metrics used in each reviewed article

Table A.6. List of metrics applied by the different reviewed articles

	Authors	Performance criteria (abbreviation)	Forecasting process end-step
1	Abe & Nakayama (2018)	R ACC MSE	III
2	Abroyan (2017)	F1	III
3	Adebiyi et al. (2014)	MSE	III
4	Adhikari & Agrawal (2013)	ACC RMSE	III
5	Agrawal et al. (2019)	ACC MSE	III
6	Aguirre et al. (2020)	CRes ARet	V
7	Akita et al. (2016)	CRes	V
8	Althelaya et al. (2018)	MAE RMSE R^2	III
9	de Araújo et al. (2018)	MSE MAPE TheilU ARV POCID EF	III
10	de Araújo et al. (2012)	MSE MAPE TheilU ARV POCID EF	III
11	Assis et al. (2018)	ACC	III
12	Baek & Kim (2018)	MSE MAPE MAE	III
13	Ballings et al. (2015)	AUC	III
14	Bao, et al. (2017)	MAPE R TheilU ARet	V
15	Bekiros (2013)	ACC PT HM MSE ARet SR	V
16	Bildirici et al. (2010)	RMSE	III
17	Börjesson & Singull (2020)	ACC MAPE	III
18	Borovkova & Tsiamas (2019)	AUC	III
19	Borovykh et al. (2018)	MASE ACC RMSE	III
20	Cao et al. (2011)	MAE MAPE MSE	III
21	Carta et al. (2021)	ACC MDD COV SorR CRes	V
22	Chakraborty (2019)	ARet	V
23	Chandra & Chand (2016)	RMSE	III
24	Chang et al. (2012)	MAPE	III
25	Chaudhari & Thakkar,2021)	DPA F1 Prec Rec	III
26	Chen et al. (2017a)	ARet SR ACC	V
27	Chen et al. (2016)	SR ARet Cres	V
28	Chen et al. (2017b)	ACC	III
29	Chen et al. (2015)	ACC	III
30	Chen et al. (2018)	RMSE MAPE ACC	III
31	Chen et al. (2020)	SR R^2 EV	V
32	Chen, Chen et al. (2013)	RMSE	III
33	Chen et al. (2013)	R RMSE	III
34	Chen & Ge (2019)	ACC ARet	V
35	Chen et al. (2018a)	ACC MAE MAPE RMSE	III
36	Chen et al. (2017c)	MAE MAPE RMSE	III
37	Chen et al. (2018b)	MSE MAE	III
38	Chong et al. (2017b)	NMSE RMSE MAE MI	III
39	Colliri & Zhao (2021)	ARet SR ACC	V
40	Dai et al. (2012)	RMSE MAE MAPE RMSPE DPA	III
41	Das et al. (2018b)	ACC Prec Rec F1 AUROC	III
42	Das et al. (2021)	MSE MAE MAPE R TheilU	III
43	Das et al. (2018a)	R	III
44	Dash & Dash (2016)	ARet	V
45	De Oliveira et al. (2013)	MAPE RMSE TheilU POCID	III
46	Deng et al. (2017)	CRes ARet SR	V
47	Deszi & Nistor (2014)	RMSE MAE	III
48	Ding et al. (2015)	ACC MCC CRes	V
49	Dingli & Fournier (2017a)	ACC RMSE	III
50	Dingli & Fournier (2017b)	ACC RMSE	III
51	Elliot & Hsu (2017)	MAE RMSE	III
52	Fan et al. (2017)	R	III
53	Feng et al. (2018a)	R RMSE	III
54	Feng et al. (2018b)	MSE R^2	III
55	Feuerriegel & Prendinger (2016)	ARet SR CR Vol	V
56	Fischer & Krauss (2018a)	ARet Vol SR ACC	V
57	Gu et al. (2020)	R^2 SR	V

(continued on next page)

Table A.6. List of metrics applied by the different reviewed articles (*continued*)

	Authors	Performance criteria (abbreviation)	Forecasting process end-step
58	Gunduz et al. (2017)	F1	III
59	Guresen et al. (2011)	MSE MAE MAPE	III
60	Han et al. (2018)	Prec Rec F1	III
61	Hansson & Nilsson (2017)	MSE ACC	III
62	Hansun & Young (2021)	RMSE MAPE	III
63	Hao & Gao (2020)	ACC	III
64	Heaton et al. (2016)	ARet	V
65	Henrique et al. (2018)	RMSE MAPE	III
66	Hernandez & Abad (2018)	ACC	III
67	Hiransha et al. (2018)	MAPE	III
68	Hsieh et al. (2011)	RMSE MAE MAPE TheilU	III
69	Huynh et al. (2017)	ACC	III
70	Iwasaki & Chen (2018)	ACC R^2	III
71	Jeong & Kim (2019)	CRes R	V
72	Ji et al. (2021)	R^2 MAE RMSE	III
73	Jiang et al. (2020)	MAE MSE MAPE	III
74	Kao et al. (2013)	RMSE MAE MAPE RMSPE DPA	III
75	Kara et al. (2011)	ACC	III
76	Karaoglu et al. (2017)	MSE RMSE MAE RSE R^2	III
77	Kelotra & Pandey (2020)	MSE RMSE	III
78	Khare et al. (2017)	RMSE	III
79	Kim & Ahn (2012)	ACC	III
80	Kraus & Feuerriegel (2017)	MSE RMSE MAE ACC AUC	III
81	Krauss et al. (2017)	ARet MDD CR	V
82	Kumar et al. (2016)	ACC	III
83	Kumar et al. (2021)	ACC MSE RMSE MAE MAPE	III
84	Labiad et al. (2018)	ACC F1 Prec Rec	III
85	Lachiheb & Gouider (2018)	ACC MSE	III
86	Lee & Soo (2018)	RMSE CRes	V
87	Lee & Yoo (2018)	ACC ARet	V
88	Li et al. (2017)	Prec Rec F1	III
89	Li & Liao (2018)	ACC F1	III
90	Li et al. (2020)	ACC F1 Prec Rec	III
91	Li et al. (2017)	MSE NRMSE MAE	III
92	Li & Tam (2018)	ACC MAPE	III
93	Liang et al. (2017)	DPA	III
94	Liao & Wang (2010)	ARet	V
95	Lien Minh et al. (2018)	NMSE RMSE MAE MI	III
96	Lim et al. (2019)	SR SorR MDD CR ACC ARet Vol APL MAR	V
97	Liu (2018)	ACC	III
98	Liu et al. (2021)	ACC	III
99	Liu et al. (2017)	ARet	V
100	Liu et al. (2020)	ACC MCC	III
101	Long et al. (2019)	ARet SR MDD Vol	V
102	Lu & Wu (2011)	RMSE MAE MAPE ACC	III
103	Lv et al. (2019)	ACC Prec Rec F1 AUC ARet CR MDD	V
104	Ma et al. (2021)	MSE MAE ACC ARet CRes Vol IR MDD	V
105	Mallikarjuna & Rao (2019)	RMSE	III
106	Martínez-Miranda et al. (2016)	CRes Vol	V
107	Matsubara et al. (2018)	ACC MCC	III
108	Mehtab & Sen (2020)	ACC Prec Rec DPA R RMSE	III
109	Minami & Minami (2018)	RMSE	III
110	Moghaddam et al. (2016)	R^2	III
111	Mohanty et al. (2021)	RMSE MAE MAPE	III
112	Mourelatos et al. (2018)	ARet Vol SR ACC	V
113	Mundra et al. (2020)	ACC MAPE info	V
114	Nabipour et al. (2020)	F1 ACC ROC AUC	III
115	Nabipour et al. (2020)	R^2 MAE MAPE	III
116	Nascimento & Cristo (2015)	MAPE RMSE	III
117	Navon & Keller (2017)	CRes	V
118	Nayak et al. (2012)	MAE	III
119	Ndikum (2020)	MSE	III
120	Ni et al. (2011)	ACC	III
121	Nikou et al. (2019)	MSE MAE MAPE	III
122	Pang et al. (2020)	ACC MSE	III
123	Parida et al. (2016)	RMSE MAPE MAE	III
124	Patil et al. (2020)	RMSE MAPE MAE	III
125	Persio & Honchar (2017)	ACC LogLoss	III
126	Porshnev et al. (2013)	ACC Prec Rec F1	III
127	Qin et al. (2017)	RMSE MAE MAPE	III
128	Qiu & Song (2016)	ACC	III
129	Qiu et al. (2020)	ACC MSE RMSE F1 Prec Rec AUC	III
130	Rasekhschaffe & Jones (2019)	ARet Vol R	V
131	Rather (2011)	MAE MSE MAPE	III
132	Rather (2014)	MSE MAE	III
133	Rather et al. (2015)	MSE MAE R	III

(*continued on next page*)

Table A.6. List of metrics applied by the different reviewed articles (*continued*)

	Authors	Performance criteria (abbreviation)	Forecasting process end-step
134	Rodríguez-González et al. (2011)	ACC	III
135	Roondiwala et al. (2017)	RMSE	III
136	Rout et al. (2017)	RMSE MAPE	III
137	Roy Choudhury et al. (2020)	R^2 RMSE MAPE	III
138	Rubesam (2019)	ARet Vol SR MDD info	V
139	Saifan et al. (2020)	CRes Vol Alpha SR SorR IR MDD	V
140	Samarawickrama & Fernando (2018)	MAE MAPE	III
141	Selvin et al. (2017)	ACC	III
142	Sezer & Ozbayoglu (2018)	ARet	V
143	Sezer et al. (2020)	Prec Rec F1 ARet	V
144	Sezer et al. (2017)	ARet ACC CRes MDD 6 info	V
145	Shen et al. (2018)	ARet	V
146	Shin et al. (2019)	CRes MDD	V
147	Shynkevich et al. (2017)	ACC WR ARet SR	V
148	Si et al. (2018)	CRes SR	V
149	Siami-Namini & Namin (2018)	RMSE	III
150	Sim et al. (2019)	HIT Rec Spec	III
151	Singh & Srivastava (2016)	SMAPE POCID MAPE RMSE ACC R	III
152	Sohangir et al. (2018)	ACC F1 AUC Prec	III
153	Song et al. (2020)	MAE RMSE MAPE	III
154	Soto et al. (2014)	RMSE	III
155	Takahashi & Chen (2017)	RMSE	III
156	Thakkar & Chaudhari (2020)	DPA	III
157	Chaudhari & Thakkar (2021)	RMSE DPA F1 Prec Rec	III
158	Ticknor (2013)	MAPE	III
159	Tran et al. (2018)	ACC Prec Rec F1	III
160	Tsantekidis et al. (2020)	Prec Rec F1	III
161	U et al. (2020)	Prec Rec F1	III
162	Vargas et al. (2018)	ACC	III
163	Vijh et al. (2020)	RMSE MAPE	III
164	Wang et al. (2011)	MAE RMSE MAPE	III
165	Wang et al. (2020)	ACC BCE ROC SAR	III
166	Wang et al. (2018)	F1 Prec Rec ACC AUC	III
167	Wen et al. (2019)	ACC Prec Rec F1	III
168	Wen et al. (2010)	CRes R^2 MSE	V
169	Wen et al. (2020)	RMSE MAPE	III
170	Wu et al. (2021)	ACC	III
171	Xu & Keselj (2019)	ACC	III
172	Yan & Ouyang (2017)	MAPE TheilU	III
173	Yan et al. (2021)	ACC	III
174	Yang et al. (2017)	ACC	III
175	Yang et al. (2020)	F1 Prec Rec ACC	III
176	Yong et al. (2017)	RMSE MAPE CRes SR	V
177	Yu et al. (2020)	ARet	V
178	Yuan et al. (2018)	MSE	III
179	Yun et al. (2020)	ACC R ARet Vol SR IR	V
180	Zhang & Maringer (2016)	SR	V
181	Zhang (2015)	MSE	III
182	Zhang et al. (2017)	MSE	III
183	Zhang et al. (2016)	ACC G-Mean	III
184	Zhang & Tan (2018)	ARet IR IC	V
185	Zhang et al. (2020)	SR SorR MDD CR ACC ARet Vol APL MAR	V
186	Zhong & Enke (2019)	MSE ARet Vol	III
187	Zhou (2019)	ARet SR	V
188	Zhou et al. (2019)	MAE RMSE MAPE	III
189	Zhou et al. (2018)	RMSRE DPA	III
190	Zhuge et al. (2017)	MSE	III

Appendix 3. Hardware and software

Hardware: In order to assess the reproducibility of the models, two different hardwares were used to perform the computation.

HARDWARE	PC1	PC2
Processor	AMD Ryzen 9 3950 X	Intel i7-8700
CPU	DDR4	DDR3
GPU	Nvidia RTX 3090 24 GB	Nvidia GTX 1060i 6 GB
RAM	64 GB	16 GB

Software: The two computers run similar software with, sometimes, marginal differences in releases.

SOFTWARE		PC1	PC2
OS		Windows 10 Pro	Windows 10 Pro
Anaconda		1.9.12	1.9.12
Spyder		4.1.5	4.1.5
Python		3.8	3.7
Cuda		11.0.221	10.2.89
CUDNN			7.6.5
Pytorch		1.7.0	1.7.0
Tensorflow		2.1.0	2.1.0
Tensorflow-GPU		2.3.0	2.3.0
Keras		2.3.1	2.3.1
scikit-learn		0.23.2	0.23.2
Stable Baselines3		1.0	1.0
Numpy		1.19.2	1.19.2
Pandas		1.1.3	1.1.3
Matplotlib		3.3.2	3.3.2
Sqlite		3.33.0	3.33.0
pyts		0.11.0	0.11.0
Neptune	Client	0.9.5	0.9.4
	Contrib	0.27.1	0.27.1

In order to secure reproducibility to the larger extent possible, we applied several strategies:

- Seed defined for Python, Numpy, TensorFlow and/or Pytorch (both CPU and GPU)
- Deterministic backend forced for CUDNN
- debug environment variable CUBLAS_WORKSPACE_CONFIG defined to ":4096:8″

Appendix 4. Deductive reasoning: Additional example

We could compare algorithms C and D, predicting respectively $\widehat{r}_{Ci}$ and $\widehat{r}_{Di}$. Let's assume that $\widehat{r}_{Di}$ is equal to $-r_i$, predicting a positive return when the return will be negative, and predicting a negative return when the actual return will be positive. Let's assume that $\widehat{r}_{Di}$ is equal to $3 * r_i$. The MSE of C and D will be equal $(r_i - \widehat{r}_{Ci})^2 = (r_i - \widehat{r}_{Di})^2$ and MAE of C and D will be equal too: $|r_i - \widehat{r}_{Ci}| = |r_i - \widehat{r}_{Di}|$. Similarly, the RMSE and MAPE of the two algorithms will be the same. But the financial results of the algorithms will be different:

- Algorithm C will trigger investments each time the return is positive and no investment when the return is negative, maximizing the loss.
- Algorithm D will generate the perfect investment strategy, with only positive daily returns and no missed opportunity, like a perfect theoretical back-trading. The return will be maximized.

References

Abe, M., & Nakayama, H. (2018). Deep learning for forecasting stock returns in the cross-section. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10937 LNAI*, 273–284. https://doi.org/10.1007/978-3-319-93034-3_22.

Abroyan, N. (2017). Neural networks for financial market risk classification. *Frontiers in Signal Processing, 1*(2), 62–66. https://doi.org/10.22606/fsp.2017.12002

Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Comparison of ARIMA and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics, 2014*. https://doi.org/10.1155/2014/614342

Adhikari, R., & Agrawal, R. K. (2013). A combination of artificial neural network and random walk models for financial time series forecasting. *Neural Computing and Applications 2013 24:6, 24*(6), 1441–1449. https://doi.org/10.1007/S00521-013-1386-Y.

Agrawal, M., Khan, A. U., & Shukla, P. K. (2019). Stock price prediction using technical indicators: A predictive model using optimal deep learning. *International Journal of Recent Technology and Engineering, 8*(2), 2297–2305. https://doi.org/10.35940/IJRTE.B3048.078219

Aguirre, A. A. A., Medina, R. A. R., & Méndez, N. D. D. (2020). Machine learning applied in the stock market through the Moving Average Convergence Divergence (MACD) indicator. *Investment Management and Financial Innovations, 17*(4), 44–60. https://doi.org/10.21511/imfi.17(4).2020.05

Akita, R., Yoshiro, A., Matsubara, T., & Uehara, K. (2016). Deep learning for stock prediction using numerical and textual information. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)* (pp. 1–6).

Althelaya, K. A., El-Alfy, E. S. M., & Mohammed, S. (2018). Evaluation of bidirectional LSTM for short and long-term stock market prediction. In *2018 9th International Conference on Information and Communication Systems*. https://doi.org/10.1109/IACS.2018.8355458

Amédée-Manesme, C. O., Barthélémy, F., & Maillard, D. (2019). Computation of the corrected Cornish-Fisher expansion using the response surface methodology: Application to VaR and CVaR. *Annals of Operations Research, 281*(1–2), 423–453. https://doi.org/10.1007/s10479-018-2792-4

Araújo, R. de A., Nedjah, N., de Seixas, J. M., Oliveira, A. L. I., & Meira, S. R. d. L. (2018). Evolutionary-morphological learning machines for high-frequency financial time series prediction. *Swarm and Evolutionary Computation*, *42*, 1–15. https://doi.org/10.1016/J.SWEVO.2018.03.009.

Araújo, R. de A., Oliveira, A. L. I., Soares, S., & Meira, S. (2012). A quantum-inspired evolutionary learning process to design dilation-erosion perceptrons for financial forecasting. *Learning and Nonlinear Models*, *10*(3), 192–201. https://doi.org/10.21528/lnlm-vol10-no3-art6.

Assis, C. A. S., Pereira, A. C. M., Carrano, E. G., Ramos, R., & Dias, W. (2018). Restricted Boltzmann machines for the prediction of trends in financial time series. *Proceedings of the International Joint Conference on Neural Networks*, *2018-July*. https://doi.org/10.1109/IJCNN.2018.8489163.

Auer, B. R., & Schuhmacher, F. (2013). Robust evidence on the similarity of Sharpe ratio and drawdown-based hedge fund performance rankings. *Journal of International Financial Markets, Institutions and Money, 24*(1), 153–165. https://doi.org/10.1016/j.intfin.2012.11.010

Baek, Y., & Kim, H. Y. (2018). ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Systems with Applications, 113*, 457–480. https://doi.org/10.1016/j.eswa.2018.07.019

Ballings, M., Van Den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications, 42* (20), 7046–7056. https://doi.org/10.1016/j.eswa.2015.05.013

Bao, W., Yue, J., Rao, Y. (2017). *A deep learning framework for financial time series using stacked autoencoders and long-short term memory.pdf*. https://doi.org/10.1371/journal.pone.0180944.

Bekiros, S. D. (2013). Irrational fads, short-term memory emulation, and asset predictability. *Review of Financial Economics, 22*(4), 213–219. https://doi.org/10.1016/J.RFE.2013.05.005

Bildirici, M., Alp, E. A., & Ersin, Ö.Ö. (2010). TAR-cointegration neural network model: An empirical analysis of exchange rates and stock returns. *Expert Systems with Applications, 37*(1), 2–11. https://doi.org/10.1016/J.ESWA.2009.07.077

Börjesson, L., & Singull, M. (2020). Forecasting financial time series through causal and dilated convolutional neural networks. *Entropy, 22*(10), 1–20. https://doi.org/10.3390/e22101094

Borovkova, S., & Tsiamas, I. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. *Journal of Forecasting, 38*(6), 600–619. https://doi.org/10.1002/for.2585

Borovykh, A., Bohte, S., & Oosterlee, C. W. (2018). Dilated convolutional neural networks for time series forecasting. *Journal of Computational Finance*. https://doi.org/10.21314/JCF.2019.358

Botchkarev, A. (2018). Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology. *ArXiv 1809.03006*, 1–37. http://arxiv.org/abs/1809.03006.

Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A systematic review. *Expert Systems with Applications, 156*(October). https://doi.org/10.1016/j.eswa.2020.113464

Cao, Q., Parry, M. E., & Leggio, K. B. (2011). The three-factor model and artificial neural networks: Predicting stock price movement in China. *Annals of Operations Research, 185*(1), 25–44. https://doi.org/10.1007/s10479-009-0618-0

Carta, S., Ferreira, A., Podda, A. S., Reforgiato Recupero, D., & Sanna, A. (2021). Multi-DQN: An ensemble of Deep Q-learning agents for stock market forecasting. *Expert Systems with Applications*, *164*(July 2020), 113820. https://doi.org/10.1016/j.eswa.2020.113820.

Chakraborty, S. (2019). Capturing Financial markets to apply Deep Reinforcement Learning. *ArXiv 1907.04373*, 1–17. http://arxiv.org/abs/1907.04373.

Chandra, R., & Chand, S. (2016). Evaluation of co-evolutionary neural network architectures for time series prediction with mobile application in finance. *Applied Soft Computing, 49*, 462–473. https://doi.org/10.1016/J.ASOC.2016.08.029

Chang, P. C., Wang, D. D., & Zhou, C. L. (2012). A novel model by evolving partially connected neural network for stock price trend forecasting. *Expert Systems with Applications, 39*(1), 611–620. https://doi.org/10.1016/J.ESWA.2011.07.051

Chaudhari, K., & Thakkar, A. (2021). iCREST: International cross-reference to exchange-based stock trend prediction using long short-term memory. *Applied Soft Computing and Communication Networks, 323–338*. https://doi.org/10.1007/978-981-33-6173-7_22

Chen, H., Xiao, K., Sun, J., & Wu, S. (2017a). A double-layer neural network framework for high-frequency forecasting. *ACM Transactions on Management Information Systems (TMIS), 7*(4). https://doi.org/10.1145/3021380

Chen, J. F., Chen, W. L., Huang, C. P., Huang, S. H., & Chen, A. P. (2017). Financial time-series data analysis using deep convolutional neural networks. *Proceedings - 2016 7th International Conference on Cloud Computing and Big Data, CCBD 2016*, 87–92. https://doi.org/10.1109/CCBD.2016.027.

Chen, J., Wu, W., & Tindall, M. L. (2016). Hedge fund return prediction and fund selection: A machine-learning approach. *Financial Industry Studies Department, Dallas Fed, November*. https://www.dallasfed.org/banking/fis/~/media/documents/banking/occasional/1604.pdf.

Chen, K., Zhou, Y., & Dai, F. (2015). A LSTM-based method for stock returns prediction: A case study of China stock market. *Proceedings – 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 2823–2824. https://doi.org/10.1109/BIGDATA.2015.7364089.

Chen, L., Qiao, Z., Wang, M., Wang, C., Du, R., & Stanley, H. E. (2018). Which artificial intelligence algorithm better predicts the Chinese stock market? *IEEE Access, 6*, 48625–48633. https://doi.org/10.1109/ACCESS.2018.2859809

Chen, L., Pelger, M., & Zhu, J. (2020). Deep learning in asset pricing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3350138

Chen, M. Y., Chen, D. R., Fan, M. H., & Huang, T. Y. (2013). International transmission of stock market movements: An adaptive neuro-fuzzy inference system for analysis of TAIEX forecasting. *Neural Computing and Applications, 23*(SUPPL1), 369–378. https://doi.org/10.1007/S00521-013-1461-4

Chen, M. Y., Fan, M. H., Chen, Y. L., & Wei, H. M. (2013). Design of experiments on neural network's parameters optimization for time series forecasting in stock markets. *Neural Network World, 23*(4), 369–393. https://doi.org/10.14311/NNW.2013.23.023

Chen, S., & Ge, L. (2019). Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction. *Quantitative Finance, 19*(9), 1507–1515. https://doi.org/10.1080/14697688.2019.1622287

Chen, W., Yeo, C. K., Lau, C. T., & Lee, B. S. (2018). Leveraging social media news to predict stock index movement using RNN-boost. *Data and Knowledge Engineering, 118*, 14–24. https://doi.org/10.1016/J.DATAK.2018.08.003

Chen, W., Zhang, Y., Yeo, C. K., Lau, C. T., & Lee, B. S. (2017). Stock market prediction using neural network through news on online social networks. *2017 International Smart Cities Conference, ISC2 2017*. https://doi.org/10.1109/ISC2.2017.8090834.

Chen, Y., Wu, J., & Bu, H. (2018). Stock Market Embedding and Prediction: A Deep Learning Method. *2018 15th International Conference on Service Systems and Service Management, ICSSSM 2018*. https://doi.org/10.1109/ICSSSM.2018.8464968.

Chong, E., Han, C., & Park, F. C. (2017a). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications, 83*(September), 187–205. https://doi.org/10.1016/j.eswa.2017.04.030

Chong, E., Han, C., & Park, F. C. (2017b). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications, 83*, 187–205. https://doi.org/10.1016/j.eswa.2017.04.030

Colliri, T., & Zhao, L. (2021). Stock market trend detection and automatic decision-making through a network-based classification model. *Natural Computing, 9*. https://doi.org/10.1007/s11047-020-09829-9

Dai, W., Wu, J. Y., & Lu, C. J. (2012). Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. *Expert Systems with Applications, 39*(4), 4444–4452. https://doi.org/10.1016/J.ESWA.2011.09.145

Das, S. R., Mokashi, K., & Culkin, R. (2018). Are markets truly efficient? Experiments using deep learning algorithms for market movement prediction. *Algorithms, 11*(9), 1–19. https://doi.org/10.3390/a11090138

Das, Sudeepa, Sahu, T. P., Janghel, R. R., & Sahu, B. K. (2021). Effective forecasting of stock market price by using extreme learning machine optimized by PSO-based group oriented crow search algorithm. In *Neural Computing and Applications* (Vol. 0123456789). Springer London. https://doi.org/10.1007/s00521-021-06403-x.

Das, S., Behera, R. K., Kumar, M., & Rath, S. K. (2018). Real-time sentiment analysis of twitter streaming data for stock prediction. *Procedia Computer Science, 132*(Iccids), 956–964. https://doi.org/10.1016/j.procs.2018.05.111

Dash, R., & Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. *The Journal of Finance and Data Science, 2*(1), 42–57. https://doi.org/10.1016/J.JFDS.2016.03.002

De Oliveira, F. A., Nobre, C. N., & Zárate, L. E. (2013). Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4, Petrobr as, Brazil. *Expert Systems with Applications, 40*(18), 7596–7606. https://doi.org/10.1016/J.ESWA.2013.06.071

Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2017). Deep direct reinforcement learning for financial signal representation and trading. *IEEE Transactions on Neural Networks and Learning Systems, 28*(3), 653–664. https://doi.org/10.1109/TNNLS.2016.2522401

Deszi, Eva; Nistor, I.A., 2014. Can deep machine learning outsmart the market? A comparison between econometric modelling nad long-short term memory. *Lincolin Arsyad*, *3*(2), 1–46. http://journal.stainkudus.ac.id/index.php/equilibrium/article/view/1268/1127.

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction Xiao. *Journal of Scientific and Industrial Research*, 2327–2333. https://www.aaai.org/ocs/index.php/IJCAI/IJCAI15/paper/view/11031/10986.

Dingli, A., & Fournier, K. S. (2017a). Financial time series forecasting - a deep learning approach. *International Journal of Machine Learning and Computing, 7*(5), 118–122. https://doi.org/10.18178/ijmlc.2017.7.5.632

Dingli, A., & Fournier, K. S. (2017b). Financial time series forecasting – A machine learning approach. *Machine Learning and Applications: An International Journal, 4*(1/2/3), 11–27. https://doi.org/10.5121/mlaij.2017.4302

Elliot, A., & Hsu, C. H. (2017). Time series prediction : Predicting Stock Price. *ArXiv 1710.05751*, *2*. http://arxiv.org/abs/1710.05751.

Fan, J., Xue, L., & Yao, J. (2017). Sufficient forecasting using factor models. *Journal of Econometrics, 201*(2), 292–306. https://doi.org/10.1016/J.JECONOM.2017.08.009

Feng, G., He, J., & Polson, N. G. (2018). Deep learning for predicting asset returns. *ArXiv, 1804*(09314), 1–23.

Feng, G., Polson, N. G., & Xu, J. (2018). Deep Learning in Characteristics-Sorted Factor Models. *ArXiv 1805.01104*, 1–41. http://arxiv.org/abs/1805.01104.

Feuerriegel, S., & Prendinger, H. (2016). News-based trading strategies. *Decision Support Systems, 90*, 65–74. https://doi.org/10.1016/J.DSS.2016.06.020

Fischer, T., & Krauss, C. (2018). Deep learning with LSTM networks for Financial Market Predictions. *European Journal of Operational Research, 270*(2), 1–34. https://www.econstor.eu/bitstream/10419/157808/1/886576210.pdf.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies, 33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Gunduz, H., Yaslan, Y., & Cataltepe, Z. (2017). Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations. *Knowledge-Based Systems, 137*, 138–148. https://doi.org/10.1016/j.knosys.2017.09.023

Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications, 38*(8), 10389–10397. https://doi.org/10.1016/J.ESWA.2011.02.068

Han, S., Hao, X., & Huang, H. (2018). An event-extraction approach for business analysis from online Chinese news. *Electronic Commerce Research and Applications, 28*, 244–260. https://doi.org/10.1016/J.ELERAP.2018.02.006

Hansson, M., & Nilsson, B. (2017). On stock return prediction with LSTM networks. *Seminar 1st of June 2017*.

Hansun, S., & Young, J. C. (2021). Predicting LQ45 financial sector indices using RNN-LSTM. *Journal of Big Data, 8*(1). https://doi.org/10.1186/s40537-021-00495-x

Hao, Y., & Gao, Q. (2020). Predicting the trend of stock market index using the hybrid neural network based on multiple time scale feature learning. *Applied Sciences (Switzerland), 10*(11). https://doi.org/10.3390/app10113961

Heaton, J. B., Polson, N., & Witte, J. (2016). Deep learning for finance: Deep portfolios. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.2838013

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *Journal of Finance and Data Science, 4*(3), 183–201. https://doi.org/10.1016/j.jfds.2018.04.003

Hernandez, J., & Abad, A. G. (2018). Learning from multivariate discrete sequential data using a restricted Boltzmann machine model. In *2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence*. https://doi.org/10.1109/COLCACI.2018.8484854

Hiransha, M., Gopalakrishnan, E., Menon, V., & Soman, K. (2018). NSE stock market prediction using deep-learning models. *Procedia Computer Science, 132*. https://doi.org/10.1016/j.procs.2018.05.050

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process, 5*(2), 01–11. https://doi.org/10.5121/ijdkp.2015.5201

Hsieh, T. J., Hsiao, H. F., & Yeh, W. C. (2011). Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied Soft Computing, 11*(2), 2510–2525. https://doi.org/10.1016/J.ASOC.2010.09.007

Huang, J., Chai, J., & Cho, S. (2020). Deep learning in finance and banking: A literature review and classification. *Frontiers of Business Research in China, 14*(1). https://doi.org/10.1186/s11782-020-00082-6

Huynh, H. D., Dang, L. M., & Duong, D. (2017). A new model for stock price movements prediction using deep neural network. *ACM International Conference Proceeding Series, 2017-December*, 57–62. https://doi.org/10.1145/3155133.3155202.

Iwasaki, H., & Chen, Y. (2018). Topic sentiment asset pricing with DNN supervised learning. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3228485

Jeong, G., & Kim, H. Y. (2019). Improving financial trading decisions using deep Q-learning: Predicting the number of shares, action strategies, and transfer learning. *Expert Systems with Applications, 117*, 125–138. https://doi.org/10.1016/J.ESWA.2018.09.036

Ji, X., Wang, J., & Yan, Z. (2021). A stock price prediction method based on deep learning technology. *International Journal of Crowd Science, 5*(1), 55–72. https://doi.org/10.1108/ijcs-05-2020-0012

Jiang, M., Jia, L., Chen, Z., & Chen, W. (2020). The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm. *Annals of Operations Research*. https://doi.org/10.1007/s10479-020-03690-w

Kao, L. J., Chiu, C. C., Lu, C. J., & Yang, J. L. (2013). Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. *Neurocomputing, 99*, 534–542. https://doi.org/10.1016/j.neucom.2012.06.037

Kara, Y., Acar Boyacioglu, M., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications, 38*(5), 5311–5319. https://doi.org/10.1016/J.ESWA.2010.10.027

Karaoglu, S., Arpaci, U., & Serkan, A. (2017). A deep learning approach for optimization of systematic signal detection in financial trading systems with big data. *International Journal of Intelligent Systems and Applications in Engineering*, 31–36. https://doi.org/10.18201/IJISAE.2017SPECIALISSUE31421

Kelotra, A., & Pandey, P. (2020). Stock market prediction using optimized deep-ConvLSTM Model. *Https://Home.Liebertpub.Com/Big, 8*(1), 5–24. https://doi.org/10.1089/BIG.2018.0143.

Khare, K., Darekar, O., Gupta, P., & Attar, V. Z. (2017). Short term stock price prediction using deep learning. *RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings, 2018-January*, 482–486. https://doi.org/10.1109/RTEICT.2017.8256643.

Kim, K. J., & Ahn, H. (2012). Simultaneous optimization of artificial neural networks for financial forecasting. *Applied Intelligence, 36*(4), 887–898. https://doi.org/10.1007/S10489-011-0303-2

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems, 104*, 38–48. https://doi.org/10.1016/j.dss.2017.10.001

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research, 259*(2), 689–702. https://doi.org/10.1016/J.EJOR.2016.10.031

Kumar, D., Meghwani, S. S., & Thakur, M. (2016). Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. *Journal of Computational Science, 17*, 1–13. https://doi.org/10.1016/J.JOCS.2016.07.006

Kumar, D., Sarangi, P. K., & Verma, R. (2021). A systematic review of stock market prediction using machine learning and statistical techniques. *Materials Today: Proceedings, xxxx*. https://doi.org/10.1016/j.matpr.2020.11.399

Labiad, B., Berrado, A., & Benabbou, L. (2018). Short term prediction framework for Moroccan stock market using artificial neural networks. *ACM International Conference Proceeding Series, 10*(1145/3289402), 3289520.

Lachiheb, O., & Gouider, M. S. (2018). A hierarchical Deep neural network design for stock returns prediction. *Procedia Computer Science, 126*, 264–272. https://doi.org/10.1016/J.PROCS.2018.07.260

Lee, C. Y., & Soo, V. W. (2018). Predict Stock Price with Financial News Based on Recurrent Convolutional Neural Networks. *Proceedings - 2017 Conference on Technologies and Applications of Artificial Intelligence, TAAI 2017*, 160–165. https://doi.org/10.1109/TAAI.2017.27.

Lee, S. Il, & Yoo, S. J. (2018). Threshold-based portfolio: the role of the threshold and its applications. *The Journal of Supercomputing 2018 76:10, 76*(10), 8040–8057. https://doi.org/10.1007/S11227-018-2577-1.

Li, J., Bu, H., & Wu, J. (2017). Sentiment-aware stock market prediction: A deep learning method. In *14th International Conference on Services Systems and Services Management*. https://doi.org/10.1109/ICSSSM.2017.7996306

Li, W., & Liao, J. (2018). A comparative study on trend forecasting approach for stock price time series. In *Proceedings of the International Conference on Anti-Counterfeiting, Security and Identification*. https://doi.org/10.1109/ICASID.2017.8285747

Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management, 57*(5), Article 102212. https://doi.org/10.1016/J.IPM.2020.102212

Li, Xiumin, Yang, L., Xue, F., & Zhou, H. Time series prediction of stock price using deep belief networks with intrinsic plasticity. *Proceedings of the 29th Chinese Control and Decision Conference, CCDC 2017*, 1237–1242. https://doi.org/10.1109/CCDC.2017.7978707.

Li, Z., & Tam, V. (2018). Combining the real-time wavelet denoising and long-short-term-memory neural network for predicting stock indexes. In *2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 – Proceedings*. https://doi.org/10.1109/SSCI.2017.8280883

Liang, Q., Rong, W., Zhang, J., Liu, J., & Xiong, Z. (2017). Restricted Boltzmann machine based stock market trend prediction. *Proceedings of the International Joint Conference on Neural Networks, 2017-May*, 1380–1387. https://doi.org/10.1109/IJCNN.2017.7966014.

Liao, Z., & Wang, J. (2010). Forecasting model of global stock index by stochastic time effective neural network. *Expert Systems with Applications, 37*(1), 834–841. https://doi.org/10.1016/J.ESWA.2009.05.086

Lien Minh, D., Sadeghi-Niaraki, A., Huy, H. D., Min, K., & Moon, H. (2018). Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. *IEEE Access, 6*, 55392–55404. https://doi.org/10.1109/ACCESS.2018.2868970

Lim, B., Zohren, S., & Roberts, S. (2019). Enhancing time-series momentum strategies using deep neural networks. *The Journal of Financial Data Science, 1*(4), 19–38. https://doi.org/10.3905/jfds.2019.1.015

Liu, H. (2018). Leveraging financial news for stock trend prediction with attention-based recurrent neural network. *ArXiv1811.06173*. https://arxiv.org/abs/1811.06173v1.

Liu, Q., Tao, Z., Tse, Y., & Wang, C. (2021). Stock market prediction with deep learning: The case of China. *Finance Research Letters, February*, 102209. https://doi.org/10.1016/j.frl.2021.102209.

Liu, S., Zhang, C., & Ma, J. (2017). CNN-LSTM neural network model for quantitative strategy analysis in stock markets. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-70096-0_21

Liu, S., Zhang, X., Wang, Y., & Feng, G. (2020). Recurrent convolutional neural kernel model for stock price movement prediction. *PLoS ONE, 15*(6). https://doi.org/10.1371/journal.pone.0234206

Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems, 164*, 163–173. https://doi.org/10.1016/j.knosys.2018.10.034

Lu, C. J., & Wu, J. Y. (2011). An efficient CMAC neural network for stock index forecasting. *Expert Systems with Applications, 38*(12), 15194–15201. https://doi.org/10.1016/J.ESWA.2011.05.082

Lv, D., Yuan, S., Li, M., & Xiang, Y. (2019). An empirical study of machine learning algorithms for stock daily trading strategy. *Mathematical Problems in Engineering, 2019*(2). https://doi.org/10.1155/2019/7816154

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications, 165*(September 2020). https://doi.org/10.1016/j.eswa.2020.113973.

Maillard, D. (2012). A user's guide to the Cornish fisher expansion. *SSRN Electronic Journal, January*, 1–19. https://doi.org/10.2139/ssrn.1997178.

Mallikarjuna, M., & Rao, R. P. (2019). Evaluation of forecasting methods from selected stock market returns. *Financial Innovation, 5*(1). https://doi.org/10.1186/s40854-019-0157-x

Marquering, W., & Verbeek, M. (2004). The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis, 39*(2), 407–429. https://doi.org/10.1017/S0022109000003136

Martínez-Miranda, E., McBurney, P., & Howard, M. J. W. (2016). Learning unfair trading: A market manipulation analysis from the reinforcement learning perspective. In *Proceedings of the 2016 IEEE Conference on Evolving and Adaptive Intelligent Systems*. https://doi.org/10.1109/EAIS.2016.7502499

Matsubara, T., Akita, R., & Uehara, K. (2018). Stock price prediction by deep neural generative model of news articles. *IEICE Transactions on Information and Systems, E101D*(4), 901–908. https://doi.org/10.1587/TRANSINF.2016IIP0016

Mehtab, S., & Sen, J. (2020). A Time Series Analysis-Based Stock Price Prediction Using Machine Learning and Deep Learning Models. *ArXiv 2004.11697*, 1–46. https://doi.org/10.13140/RG.2.2.14022.22085/2.

Meng, T. L., & Khushi, M. (2019). Reinforcement learning in financial markets. *Data, 4*(3), 1–17. https://doi.org/10.3390/data4030110

Minami, S., & Minami, S. (2018). Predicting equity price with corporate action events using LSTM-RNN. *Journal of Mathematical Finance, 8*(1), 58–63. https://doi.org/10.4236/JMF.2018.81005

Moghaddam, A. H., Moghaddam, M. H., & Esfandyari, M. (2016). Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science, 21*(41), 89–93. https://doi.org/10.1016/j.jefas.2016.07.002

Mohanty, D. K., Parida, A. K., & Khuntia, S. S. (2021). Financial market prediction under deep learning framework using auto encoder and kernel extreme learning machine. *Applied Soft Computing, 99*, Article 106898. https://doi.org/10.1016/j.asoc.2020.106898

Mourelatos, M., Alexakos, C., Amorgianiotis, T., & Likothanassis, S. (2018). Financial Indices Modelling and Trading utilizing Deep Learning Techniques: The ATHENS SE FTSE/ASE Large Cap Use Case. *2018 IEEE (SMC) International Conference on Innovations in Intelligent Systems and Applications, INISTA 2018*. https://doi.org/10.1109/INISTA.2018.8466286.

Mundra, A., Mundra, S., Verma, V. K., & Srivastava, J. S. (2020). A deep learning based hybrid framework for stock price prediction. *Journal of Intelligent and Fuzzy Systems, 38*(5), 5949–5956. https://doi.org/10.3233/JIFS-179681

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E., & Shahab, S. (2020). Deep learning for stock market prediction. *Entropy, 22*(8). https://doi.org/10.3390/E22080840

Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020). Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; A comparative analysis. *IEEE Access, 8*, 150199–150212. https://doi.org/10.1109/ACCESS.2020.3015966

Nascimento, J. B., & Cristo, M. (2015). The impact of structured event embeddings on scalable stock forecasting models. In *WebMedia 2015 - Proceedings of the 21st Brazilian Symposium on Multimedia and the Web* (pp. 121–124). https://doi.org/10.1145/2820426.2820467

Navon, A., & Keller, Y. (2017). Financial Time Series Prediction Using Deep Learning. *ArXiv 1711.04174*. https://arxiv.org/abs/1711.04174v1.

Nayak, S. C., Misra, B. B., & Behera, H. S. (2012). Index prediction with neuro-genetic hybrid network: A comparative analysis of performance. *2012 International Conference on Computing, Communication and Applications, ICCCA 2012*. https://doi.org/10.1109/ICCCA.2012.6179215.

Ndikum, P. (2020). Machine Learning Algorithms for Financial Asset Price Forecasting. *ArXiv 2004.01504v1*, 1–16. http://arxiv.org/abs/2004.01504.

Ni, L. P., Ni, Z. W., & Gao, Y. Z. (2011). Stock trend prediction based on fractal feature selection and support vector machine. *Expert Systems with Applications, 38*(5), 5569–5576. https://doi.org/10.1016/j.eswa.2010.10.079

Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management, 26*(4), 164–174. https://doi.org/10.1002/isaf.1459

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. In *Artificial Intelligence Review* (Vol. 53, Issue 4). Springer Netherlands. https://doi.org/10.1007/s10462-019-09754-z.

Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications : A survey. *ArXiv*.

Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2018). An innovative neural network approach for stock market prediction. *Journal of Supercomputing, 76*(3), 2098–2118. https://doi.org/10.1007/s11227-017-2228-y

Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2020). An innovative neural network approach for stock market prediction. *Journal of Supercomputing, 76*(3), 2098–2118. https://doi.org/10.1007/s11227-017-2228-y

Parida, A. K., Bisoi, R., & Dash, P. K. (2016). Chebyshev polynomial functions based locally recurrent neuro-fuzzy information system for prediction of financial and energy market data. *The Journal of Finance and Data Science, 2*(3), 202–223. https://doi.org/10.1016/J.JFDS.2016.10.001

Patil, P., Wu, C. S. M., Potika, K., & Orang, M. (2020). Stock market prediction using ensemble of graph theory, machine learning and deep learning models. *ACM International Conference Proceeding Series, 85–92*. https://doi.org/10.1145/3378936.3378972

Persio, L. D., & Honchar, O. (2017). Recurrent neural networks approach to the financial forecast of Google assets. *International Journal of Mathematics and Computers in Simulation, 11*, 7–13.

Porshnev, A., Redkin, I., & Shevchenko, A. (2013). Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis. In *Proceedings – IEEE 13th International Conference on Data Mining Workshops*. https://doi.org/10.1109/ICDMW.2013.111

Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G., & Cottrell, G. W. (2017). A dual-stage attention-based recurrent neural network for time series prediction. *IJCAI International Joint Conference on Artificial Intelligence*, 2627–2633. https://doi.org/10.24963/ijcai.2017/366

Qiu, M., & Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PLoS ONE, 11*(5), Article e0155133. https://doi.org/10.1371/JOURNAL.PONE.0155133

Qiu, Y., Yang, H.-Y., Lu, S., & Chen, W. (2020). A novel hybrid model based on recurrent neural networks for stock market timing. *Soft Computing 2020 24:20*, *24*(20), 15273–15290. https://doi.org/10.1007/S00500-020-04862-3.

Rasekhschaffe, K. C., & Jones, R. C. (2019). Machine learning for stock selection. *Financial Analysts Journal, 75*(3), 70–88. https://doi.org/10.1080/0015198X.2019.1596678

Rather, A. M. (2011). A prediction based approach for stock returns using autoregressive neural networks. In *Proceedings of the 2011 World Congress on Information and Communication Technologies, WICT 2011*. https://doi.org/10.1109/WICT.2011.6141431

Rather, A. M. (2014). A hybrid intelligent method of predicting stock returns. *Advances in Artificial Neural Systems, 2014*, 1–7. https://doi.org/10.1155/2014/246487

Rather, A. M., Agarwal, A., & Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications, 42*(6), 3234–3241. https://doi.org/10.1016/J.ESWA.2014.12.003

Rockafellar, R. T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance, 26*(7), 1443–1471. https://doi.org/10.1016/S0378-4266(02)00271-6

Rodríguez-González, A., García-Crespo, Á., Colomo-Palacios, R., Guldrís Iglesias, F., & Gómez-Berbís, J. M. (2011). CAST: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator. *Expert Systems with Applications, 38*(9), 11489–11500. https://doi.org/10.1016/J.ESWA.2011.03.023

Roondiwala, M., Patel, H., & Varma, S. (2017). Predicting stock prices using LSTM. *International Journal of Science and Research (IJSR)*. https://doi.org/10.21275/ART20172755

Rout, A. K., Dash, P. K., Dash, R., & Bisoi, R. (2017). Forecasting financial time series using a low complexity recurrent neural network and evolutionary learning approach. *Journal of King Saud University – Computer and Information Sciences, 29*(4), 536–552. https://doi.org/10.1016/J.JKSUCI.2015.06.002

Roy Choudhury, A., Abrishami, S., Turek, M., & Kumar, P. (2020). Enhancing profit from stock transactions using neural networks. *AI Communications, 33*(2), 75–92. https://doi.org/10.3233/AIC-200629

Rubesam, A. (2019). Machine learning portfolios with equal risk contributions. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3432760

Saifan, R., Sharif, K., Abu-Ghazaleh, M., & Abdel-Majeed, M. (2020). Investigating algorithmic stock market trading using ensemble machine learning methods. *Informatica (Slovenia), 44*(3), 311–325. https://doi.org/10.31449/INF.V44I3.2904

Samarawickrama, A. J. P., & Fernando, T. G. I. (2018). A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market. In *2017 IEEE International Conference on Industrial and Information Systems*. https://doi.org/10.1109/ICIINFS.2017.8300345

Sarykalin, S., Serraino, G., & Uryasev, S. (2008). Value-at-risk vs. conditional value-at-risk in risk management and optimization. *State-of-the-Art Decision-Making Tools in the Information-Intensive Age*, 270–294. https://doi.org/10.1287/educ.1080.0052.

Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017). Stock price prediction using LSTM, RNN and CNN-sliding window model. In *2017 International conference on advances in computing, communications and informatics*. https://doi.org/10.1109/ICACCI.2017.8126078

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing Journal*, *90*(May). https://doi.org/10.1016/j.asoc.2020.106181.

Sezer, O. B., & Ozbayoglu, A. M. (2018). Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Applied Soft Computing Journal, 70*(April), 525–538. https://doi.org/10.1016/j.asoc.2018.04.024

Sezer, O. B., Ozbayoglu, M., & Dogdu, E. (2017). A deep neural-network based stock trading system based on evolutionary optimized technical analysis parameters. *Procedia Computer Science, 114*, 473–480. https://doi.org/10.1016/J.PROCS.2017.09.031

Shen, G., Tan, Q., Zhang, H., Zeng, P., & Xu, J. (2018). Deep learning with gated recurrent unit networks for financial sequence predictions. *Procedia Computer Science, 131*, 895–903. https://doi.org/10.1016/J.PROCS.2018.04.298

Shin, W., Bu, S.-J., & Cho, S.-B. (2019). *Automatic Financial Trading Agent for Low-risk Portfolio Management using Deep Reinforcement Learning*. http://arxiv.org/abs/1909.03278.

Shynkevich, Y., McGinnity, T. M., Coleman, S. A., Belatreche, A., & Li, Y. (2017). Forecasting price movements using technical indicators: Investigating the impact of varying input window length. *Neurocomputing, 264*, 71–88. https://doi.org/10.1016/J.NEUCOM.2016.11.095

Si, W., Li, J., Rao, R., & Ding, P. (2018). A multi-objective deep reinforcement learning approach for stock index futures's intraday trading. *Proceedings - 2017 10th International Symposium on Computational Intelligence and Design, ISCID 2017*, *2*, 431–436. https://doi.org/10.1109/ISCID.2017.210.

Siami-Namini, Sima; Namin, A. S. (2018). *Forecasting Economics and Financial Time Series: ARIMA vs. LSTM*. ArXiv:1803.06386. https://www.researchgate.net/publication/323867492_Forecasting_Economics_and_Financial_Time_Series_ARIMA_vs_LSTM.

Sim, H. S., Kim, H. I., & Ahn, J. J. (2019). Is deep learning for image recognition applicable to stock market prediction? *Complexity, 2019*. https://doi.org/10.1155/2019/4324878

Singh, R., & Srivastava, S. (2016). Stock prediction using deep learning. *Multimedia Tools and Applications 2016 76:18*, *76*(18), 18569–18584. https://doi.org/10.1007/S11042-016-4159-7.

Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftaar, T. M. (2018). Big Data: Deep Learning for financial sentiment analysis. *Journal of Big Data, 5*(1). https://doi.org/10.1186/s40537-017-0111-6

Song, D., Busogi, M., Chung Baek, A. M., & Kim, N. (2020). Forecasting stock market index based on pattern-driven long short-term memory. *Economic Computation and Economic Cybernetics Studies and Research, 54*(3), 25–41. https://doi.org/10.24818/18423264/54.3.20.02

Soto, J., Melin, P., & Castillo, O. (2014). Optimization of interval type-2 fuzzy integrators in ensembles of ANFIS models for prediction of the Mackey-Glass time series. In *2014 IEEE Conference on Norbert Wiener in the 21st Century: Driving Technology's Future, 21CW 2014 - Incorporating the Proceedings of the 2014 North American Fuzzy Information Processing Society Conference, NAFIPS 2014, Conference Proceedings*. https://doi.org/10.1109/NORBERT.2014.6893880

Takahashi, S., & Chen, Y. (2017). Long memory and predictability in financial markets. In *The 31st Annual Conference of the Japanese Society for Artificial Intelligence* (pp. 1–3).

Thakkar, A., & Chaudhari, K. (2020). CREST: Cross-reference to exchange-based stock trend prediction using long short-term memory. *Procedia Computer Science, 167*, 616–625. https://doi.org/10.1016/J.PROCS.2020.03.328

Thakkar, A., & Chaudhari, K. (2021). A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions. *Expert Systems with Applications, 177*(March), Article 114800. https://doi.org/10.1016/j.eswa.2021.114800

Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications, 40*(14), 5501–5506. https://doi.org/10.1016/J.ESWA.2013.04.013

Tran, D. T., Magris, M., Kanniainen, J., Gabbouj, M., & Iosifidis, A. (2018). Tensor representation in high-frequency financial data for price change prediction. In *2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 – Proceedings, 2018-Janua(December 2019)*. https://doi.org/10.1109/SSCI.2017.8280812

Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., & Iosifidis, A. (2020). Using Deep Learning for price prediction by exploiting stationary limit order book features. *Applied Soft Computing Journal, 93*(October). https://doi.org/10.1016/j.asoc.2020.106401

U. JuHyok, L. Peng Yu, K. Chung Song, et al. (2020). A new LSTM based reversal point prediction method using upward/downward reversal point feature sets. *Chaos, Solitons and Fractals*, *132*, 109559. https://doi.org/10.1016/j.chaos.2019.109559.

Vargas, M. R., Dos Anjos, C. E. M., Bichara, G. L. G., & Evsukoff, A. G. (2018). Deep leaming for stock market prediction using technical indicators and financial news articles. *Proceedings of the International Joint Conference on Neural Networks, 2018-July*. https://doi.org/10.1109/IJCNN.2018.8489208.

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia Computer Science, 167*(2019), 599–606. https://doi.org/10.1016/j.procs.2020.03.326

Wang, J. Z., Wang, J. J., Zhang, Z. G., & Guo, S. P. (2011). Forecasting stock indices with back propagation neural network. *Expert Systems with Applications, 38*(11), 14346–14355. https://doi.org/10.1016/J.ESWA.2011.04.222

Wang, P., Zong, L., & Ma, Y. (2020). An integrated early warning system for stock market turbulence. *Expert Systems with Applications, 153*. https://doi.org/10.1016/j.eswa.2020.113463

Wang, Q., Xu, W., & Zheng, H. (2018). Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles. *Neurocomputing, 299*, 51–61. https://doi.org/10.1016/J.NEUCOM.2018.02.095

Wen, M., Li, P., Zhang, L., & Chen, Y. (2019). Stock market trend prediction using high-order information of time series. *IEEE Access, 7*, 28299–28308. https://doi.org/10.1109/ACCESS.2019.2901842

Wen, Q., Yang, Z., Song, Y., & Jia, P. (2010). Automatic stock decision support system based on box theory and SVM algorithm. *Expert Systems with Applications, 37*(2), 1015–1022. https://doi.org/10.1016/j.eswa.2009.05.093

Wen, Y., Lin, P., & Nie, X. (2020). Research of stock price prediction based on PCA-LSTM model. *IOP Conference Series: Materials Science and Engineering, 790*(1). https://doi.org/10.1088/1757-899X/790/1/012109

Wu, J. M. T., Li, Z., Herencsar, N., Vo, B., & Lin, J. C. W. (2021). A graph-based CNN-LSTM stock price prediction algorithm with leading indicators. *Multimedia Systems*. https://doi.org/10.1007/s00530-021-00758-w

Xu, Y., & Keselj, V. (2019). Stock Prediction using Deep Learning and Sentiment Analysis. *Proceedings – 2019 IEEE International Conference on Big Data, Big Data 2019*, 5573–5580. https://doi.org/10.1109/BIGDATA47090.2019.9006342.

Yan, H., & Ouyang, H. (2017). Financial time series prediction based on deep learning. *Wireless Personal Communications 2017 102:2, 102*(2), 683–700. https://doi.org/10.1007/S11277-017-5086-2.

Yan, X., Weihan, W., & Chang, M. (2021). Research on financial assets transaction prediction model based on LSTM neural network. *Neural Computing and Applications, 33*(1), 257–270. https://doi.org/10.1007/s00521-020-04992-7

Yang, B., Gong, Z. J., & Yang, W. (2017). Stock market index prediction using deep neural network ensemble. *Chinese Control Conference, CCC*, 3882–3887. https://doi.org/10.23919/CHICC.2017.8027964

Yang, C., Zhai, J., Tao, G., & Haajek, P. (2020). Deep learning for price movement prediction using convolutional neural network and long short-term memory. *Mathematical Problems in Engineering, 2020*. https://doi.org/10.1155/2020/2746845

Yong, B. X., Abdul Rahim, M. R., & Abdullah, A. S. (2017). A stock market trading system using deep neural network. *Communications in Computer and Information Science, 751*, 356–364. https://doi.org/10.1007/978-981-10-6463-0_31

Yu, J. R., Paul Chiou, W. J., Lee, W. Y., & Lin, S. J. (2020). Portfolio models with return forecasting and transaction costs. *International Review of Economics and Finance, 66* (November 2019), 118–130. https://doi.org/10.1016/j.iref.2019.11.002.

Yuan, Z., Zhang, R., & Shao, X. (2018). Deep and wide neural networks on multiple sets of temporal data with correlation. *ACM International Conference Proceeding Series, Part, F137704*, 39–43. https://doi.org/10.1145/3219788.3219793

Yun, H., Lee, M., Kang, Y. S., & Seok, J. (2020). Portfolio management via two-stage deep learning with a joint cost. *Expert Systems with Applications, 143*, Article 113041. https://doi.org/10.1016/j.eswa.2019.113041

Zhang, J., & Maringer, D. (2016). Using a genetic algorithm to improve recurrent reinforcement learning for equity trading. *Computational Economics, 47*(4), 551–567. https://doi.org/10.1007/s10614-015-9490-y

Zhang, L., Aggarwal, C., & Qi, G.-J. (2017). Stock price prediction via discovering multi-frequency trading patterns. *Kdd, 10*(1145/3097983), 3098117.

Zhang, L. M. (2015). Genetic deep neural networks using different activation functions for financial data mining. *Proceedings – 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 2849–2851. https://doi.org/10.1109/BIGDATA.2015.7364099.

Zhang, Xiao Dan, Li, A., & Pan, R. (2016). Stock trend prediction based on a new status box method and AdaBoost probabilistic support vector machine. *Applied Soft Computing, 49*, 385–398. https://doi.org/10.1016/J.ASOC.2016.08.026.

Zhang, X., & Tan, Y. (2018). Deep stock ranker: A LSTM neural network model for stock selection. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-93803-5_58

Zhang, Z., Zohren, S., & Roberts, S. (2020). Deep reinforcement learning for trading. *The Journal of Financial Data Science, 2*(2), 25–40. https://doi.org/10.3905/jfds.2020.1.030

Zhong, X., & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation, 5*(1). https://doi.org/10.1186/s40854-019-0138-0

Zhou, B. (2019). Deep learning and the cross-section of stock returns: Neural networks combining price and fundamental information. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3179281

Zhou, F., Zhou, H. Min, Yang, Z., & Yang, L. (2019). EMD2FNN: A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction. *Expert Systems with Applications, 115*, 136–151. https://doi.org/10.1016/J.ESWA.2018.07.065.

Zhou, X., Pan, Z., Hu, G., Tang, S., & Zhao, C. (2018). Stock market prediction on high-frequency data using generative adversarial nets. *Mathematical Problems in Engineering, 2018*. https://doi.org/10.1155/2018/4907423

Zhuge, Q., Xu, L., & Zhang, G. (2017). LSTM neural network with emotional analysis for prediction of stock price. *Engineering Letters, 25*(2). https://www.researchgate.net/publication/317750904_LSTM_neural_network_with_emotional_analysis_for_prediction_of_stock_price.