Heliyon 9 (2023) e16155

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Optimization of investment strategies through machine learning

Jiaqi Li [a], Xiaoyan Wang [b], Saleem Ahmad [c,*], Xiaobing Huang [d], Yousaf Ali Khan [e]

[a] UNSW Business School, University of New South Wales (UNSW Sydney), Sydney, 2052, NSW, Australia
[b] Accounting Department, Hebei Vocational University of Technology and Engineering, Xingtai, 054000, Hebei Province, China
[c] School of Business, Guangdong University of Foreign Studies, Guangzhou, China
[d] School of Economics, Gannan Normal University, Ganzhou, China
[e] Department of Mathematics and Statistics, Hazara University Mansehra, Pakistan

ARTICLE INFO

Keywords:
Algorithmic trading
Economic value-added strategy
Long-short term memory
Machine learning
Moving average convergence
Quantitative stock investment
Stochastic indicators

ABSTRACT

The main objective of this research is to develop a sustainable stock quantitative investing model based on Machine Learning and Economic Value-Added techniques for optimizing investment strategies. Quantitative stock selection and algorithmic trading are the two features of the model. Principal component analysis and economic value-added criteria are used in quantitative stock model for efficiently stocks selection, which may repeatedly select valuable stocks. Machine learning techniques such as Moving Average Convergence, Stochastic Indicators and Long-Short Term Memory are used in algorithmic trading. One of the first attempts, the Economic Value-Added indicators are used to appraise stocks in this study. Furthermore, the application of EVA in stock selection is exposed. Illustration of the proposed model has been done on United States stock market and finding shows that Long-Short Term Memory (LSTM) networks can more accurately forecast future stock values. The proposed strategy is feasible in all market situations, with a return that is significantly larger than the market return. As a result, the proposed approach can not only assist the market in returning to rational investing, but also assist investors in obtaining significant returns that are both realistic and valuable.

1. Introduction

Quantitative investing is gaining attraction among individual and institutional investors in the stock market, thanks to the rapid collection of financial big data and the constant improvement of machine learning algorithms [1]. Quantitative investment is a type of investment that combines knowledge of mathematical statistics, computers, and finance to generate trading models and strategies that look for profitable investment opportunities in the market and, in the end, achieve the goal of rational investment and maximum returns [2,3]. Quantitative stock selection and computerized trading are the two primary components of quantitative investment (algorithmic trading). Quantitative stock selection is the process of building a high-quality stock portfolio, utilizing an appropriate stock selection index system and quantitative statistical stock instruments analysis to achieve a return higher than the benchmark [3]. "Algorithmic trading" is defined as "*a way of executing orders using computerized pre-programmed trading instructions that account for variables such as time, price, and volume*," with the portfolio given by quantitative stock selection serving as the foundation [4–6]. Quantitative investment now accounts for a significant share of market securities transactions. As a result, the development of a long-term and beneficial quantitative investing model has become a hot issue among researchers and stock market participants [5].

* Corresponding author.
E-mail addresses: saleem.marral@gdufs.edu.cn (S. Ahmad), yousaf_hu@yahoo.com (Y.A. Khan).

https://doi.org/10.1016/j.heliyon.2023.e16155
Received 23 February 2023; Received in revised form 6 May 2023; Accepted 8 May 2023
Available online 11 May 2023
2405-8440/

Stock selection is an important foundation for quantitative investing, must not only assist investors in obtaining excess returns, but also, more crucially, assist the stock market in moving toward rational investment. Most worldwide stock markets are currently extensive with speculative and blind behaviors, which risk market stability and obstruct the stock market's primary function, namely the proper allocation of capital. As a result, the useful quantitative stock selection methodology must be able to accurately measure the enterprise's worth and rank the stocks using the scoring methodology [6,7]. In the 1950s, the first representative quantitative stock selection study was conducted. Markowitz's "mean variance" model [8] has become a new benchmark in current portfolio theory. Researcher in Ref. [9] developed the Capital Asset Pricing Model (CAPM), which advanced current finance theory by building on Markowitz's theory. At the moment, most conventional multi-factor models are based on Ross's APT model [10] and Fama-three-factor French's model notion [11], although researchers are still in search of novel component composition and construction procedures [12]. One of the most often employed stock selection models in contemporary research is multi-factor stock selection [5]. The formulation of a factor-pool and a scoring model are the most important aspects of a multi-factor stock selection model. The factors in the factor-pool, on the one hand, are highly associated with stock prices or returns. Technical and essentially indicators were the key factors evaluated in the past. The former primarily indicates the stock's market status, such as its opening and closing prices [13,14]. The latter, which includes Return on Equity (ROE), Return on Assets (ROA), and other metrics, accurately indicates the nature of the company, its financial state, and its stock's competitiveness [15–18]. These indicators, on the other hand, either ignore the company's inherent worth or only scrutinize the relative value while ignoring the underlying absolute worth. The integral value indication has not been examined in previous studies which will be considered in the underline research.

What kind of indication should be used to build the factor-pool that will be used to determine the stock's inherent value? [19] proposed EVA, which appears to be more effective indications. According to Ref. [8] EVA is an economic profit metric that helps inspire managers to increase business value. According to Ref. [20], organizations with stronger shareholder rights have a greater rate of value growth and lower capital expenditure, implying that they are more competitively sustainable. When it comes to understanding stock returns, EVA outperforms accounting measures [7,19,8]. That is why, in this study, EVA-related variables will be added to the factor pool as a stock evaluation indicator.

The scoring model, on the other hand, should be able to truly evaluate the companies' worthy of investment as the foundation of the multi-factor stock selection model [9]. Academically, the scoring methodology is likewise an excited topic. Fama, Lakonishok, and Song [10,16,21] developed a linear model of stock surplus return, claiming that the current stock price, book value of equity, and profits per share can all be explained by the current stock price, book value of equity, and profits per share. The most popular scoring model before machine learning is the linear model [5]. Machine learning is better at understanding nonlinear patterns, which are common in the stock market [22,23,24]. Principle component analysis (PCA), as a traditional machine learning technique, can essentially avoid the drawbacks of the linear model technique. It is a quantitative problem analysis that may include many stock elements into the composition, lower the data's dimension, and keep the most significant qualities of the original data [25,24]. PCA has mostly been applied to deal with attributes in previous research; however, only a few research has attempted to employ PCA to create a stock rating model. This article aims to construct a scoring model for stocks based on the primary components produced using PCA, which will be dynamically updated every year for different industries to respond to stock market fluctuations. Algorithmic trading is a vital connection for investors looking for higher profits. The benefit of algorithmic trading is that complex mathematical models take the role of human subjective assessments, and modern computer technology is used to assist investors in making trading decisions automatically. This accurately prevents investors' subjective emotions from influencing their investing decisions. In the past, algorithmic trading depended deeply on technical indicators like the Moving Average Convergence (MACD) and the Stochastic Indicator (KDJ), which were based on mathematical statistical approaches and employed sophisticated calculation formulae to identify the trend of quantitative analysis. These strategies can encourage investors in making decisions, but they can also lead to errors. Machine learning has introduced new ways to algorithm trading, allowing the detection of stock price trends and future information [18,26,27, 28]. Individuals employ several prediction algorithms [29] to investigate the future structures of the stock market. Artificial neural networks (ANNs) [30,31,32] and support vector machines (SVMs) [2,33] are two examples of modern stock market prediction models.

The findings of the study reveal that these models perform well in several aspects of stock prediction. Support Vector Machine (SVM) is a machine learning approach based on statistical learning theory. Many researchers employed support vector machine models in the field of stock prediction because of their specific benefits in undertaking small samples, high-dimensional data, and nonlinear issues. Authors of [34] employed SVM directly for stock forecasting and outlined that this technique is superior to established approaches through testing. Similarly [35], outlined a multi-core support vector machine that included global and local aspects of input data and used it to forecast the trend of stock market. Traditional SVM is more commonly employed to handle classification problems, however it is ineffective in tackling regression problems in stock price prediction. This weakness is addressed by the support vector regression machine (SVRM), which is based on nonlinear regression technology and developed on the basis of SVM. Furthermore [9] developed a genetic algorithm-enhanced SVR stock selection model based on Taiwan stock market data, which has a greater accuracy in stock price prediction. The initiation of neural networks has given rise to a novel model for stock price prediction, which is a popular topic in artificial intelligence research. Authors in Ref. [20] was the first to use neural networks to estimate the daily return of IBM common stocks in the 1990s. Through studies [1], examined the effect of multiple-factor forward artificial neural networks on Nasdaq stock price prediction. Deep learning has been established on the basis of artificial neural networks as a result of the collection of huge data in recent years and the growth of computer computing power. Stock market forecasts have also been made using deep learning models like CNN and LSTM [36,26]. Researchers of [11] constructed stock prediction models using LSTM and CNN and developed trading strategies based on the predictions' outcomes. Similarly [37], used a combination of LSTM and CNN models to forecast stock data from two perspectives: time series and stock prints. Deep learning has a number of advantages over typical machine learning approaches, one of which being the absence of feature selection [37]. These numerous machine learning systems each have their own

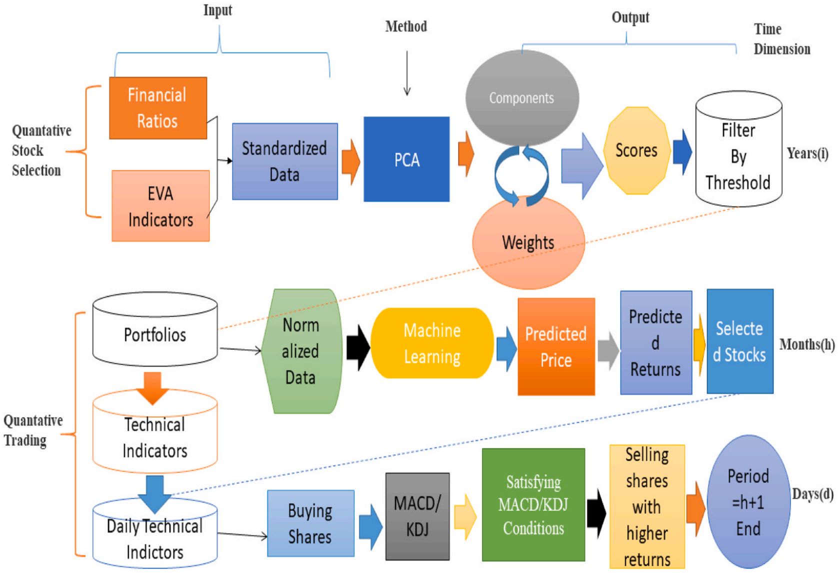

Fig. 1. Overall framework of the proposed quantitative stock investment model.

set of benefits and perform well in stock price forecasting, although, most past research has been restricted to short-term stock projections. Because the stock market is influenced by a variety of factors, only few researchers forecast the stock price a month ahead. The supremacy of each model is examined in this article, which combines SVM, ANN, LSTM, and technical indicators to estimate the price one month ahead. The medium and short-term trading of stocks is further appraised in algorithmic trading based on the expected results, and MACD and KDJ are added as stop loss indicators. Previous research has raised some concerns about this. In conclusion, past research has done a lot of work on quantitative stock selection and stock price prediction, but there is still a lot of room for improvement. To begin with, traditional factors are still used in the factor selection process, rather than indications that really measure the enterprise's underlying worth. Secondly, scoring is limited to a set factor score when constructing the scoring model and dynamic modification is not possible. Finally, stock price forecasts are still restricted to 1–5 days, and technical indicators are not completely utilized in the trading process. This article aims to address these shortcomings of the scoring model in details in the subsequent sections.

The remaining of this research is structured as follows: Section 2 present development of novel quantative investing approach based on EVA and machine learning in details, illustration of the proposed strategy on real-life data and parameter evaluation of the model is presented in section 3, detailed discussion on results are provided in section 4, while section 5 concludes this research with recommendations and future research.

2. Methodology

2.1. Formulation of the quantitative stock investment model

This study proposes a new quantitative stock investment model that employs EVA to aid in the selection of valuable stocks and incorporates machine learning technology to capture the stock market's future characteristics. Stock selection and algorithmic trading are the two primary components of the new investing model. Fig. 1 portrays the whole model framework.

Steps involves in the formulation of the proposed model is as follows.

Step 1. Introduce EVA as a criterion for determining a company's internal value.
Step II. Use PCA to dynamically update the score model.
Step II. Use machine learning technology to make monthly projections and integrate them with conventional technical indicators to trade dynamically on the basis of quantitative stock selection, ensuring the quantitative investment model's maximum return. The model's final five-year back test results demonstrate that it is stable and generates significant returns, which is of reference relevance. This is a long-term and reliable quantitative investing strategy.

2.1.1. Stock selection

Stock selection scoring is based on the factor-pool's structure. To acquire the best results, standardize the stock financial indicators and the EVA indicators $Z_{i,m,n}$, which refers to the nth indicator of the stock $\boldsymbol{m}$ in the year $\boldsymbol{i}$. Then input these indicators into the PCA model to get the score of the stock (i.e. $S_{i,m}$) based on the principal component $C_{i,m,j}$ and the corresponding weight $W_{i,j}$ of the model

Table 1
Indicators used in the quantitative stock selection.

Category	Indicators	Category	Indicators
Solvency	Current ratio	Cash flow	Net cash content of net profit
Solvency	Quick ratio	Cash flow	Company cash flow
Solvency	Equity ratio	Cash flow	Total cash recovery rate
Solvency	Long-term gearing ratio	Risk level	Capital preservation and appreciation rate
Ratio structure	Proportion of net profit attributable to the parent company	Risk level	Capital accumulation rate
Ratio structure	Working capital ratio	Development ability	Sustainable growth rate
Operating ability	Accounts Receivable Turnover Rate	Development ability	Operating income growth rate
Operating ability	Inventory turnover rate	Development ability	Capital preservation and appreciation rate
Operating ability	Cash and cash equivalent turnover rate	Development ability	Capital accumulation rate
Profitability	Return on assets	Per share indicator	Earnings per share
Profitability	Net profit margin	Per share indicator	Operating profit per share
Profitability	Return on net assets	Dividend distribution	Dividend rate
Profitability	Long-term return on capital	Dividend distribution	Earnings retention rate
Profitability	Operating profit margin	Relative value	P/E ratio
EVA indicators	EVA per share	Relative value	P/B ratio
EVA indicators	EVA growth rate	Relative value	Book-to-market ratio

EVA = Net operating profit after tax (***NOPAT***) - Total capital.

Table 2
Economic Value-Added growth rate conversion criteria.

Value	10	5	1	−1	5
Criteria	Above 100%	50%-100%	0-50%	0 or less	−100% or less

output, where $S_{i,m}$ is the score of the mth stock in the ith year. Finally, the score is enlarged to [0,100], high-value stocks (corresponding to stocks with good financial status and value, namely $S_{i,} > S_{trho}$, where S_{trho} is determined by the investor) are selected to form the stock portfolio P_i for the ith year.

2.1.2. Algorithmic trading

Algorithmic trading is based on a quantitative stock selection portfolio. There are two steps to the trading process. The first is the prediction stage. The technical indicators of all stocks in the stock portfolio for this (current) month are collected at the end of the month. T_h, ($h = 1,2 \ldots.12$), and converted to $X_{h,m,t}$ by using normalization as the machine learning model's input. Because there have been few studies in the past that have employed machine learning for monthly predictions, this article compares three machine learning models: ANN, LSTM, and SVM. The closing price of the stock m next month is y_{h+1}, which is finally obtained through machine learning prediction. The predicted stock price $y_{h+1,}$ is converted to the predicted stock return $R_{h+1} = (y_{h+1} - y_h)/y_h$. Due to the need, to maximize returns, stocks with predicted returns greater than r_{trho} are selected. Buying the selected stocks at the beginning of month $h+1$. In the month $h+1$, judge whether to sell the stock according to the MACD/KDJ indicators every day. If the indicators judge to sell stock m and the return rate reaches r_{trh}, then selling stock m, and will be cleared at the end of the month, looping into the next month.

2.2. Quantitative stock selection

This paper screened out 10 categories and 30 financial indicators based on existing literature and study [17,38,39,40,41,24]. Furthermore, two EVA indicators are used: EVA added value (see Eqn. (1)) and EVA growth rate, because [42] said that if the company's EVA performed better, it would have a brighter future. Solvency, ratio structure, operational ability, profitability, cash flow analysis, risk level, development ability, per share indicator, relative value, dividend distribution, and EVA indicators are the 11 categories of indicators. Table 1 shows the substance of the specific indicators.

(***IC***)*Weighted average cost of capital (***WACC***) (1)

It is important to note that while the calculating technique employed in this article is fixed, the computed growth rate varies widely; the EVA growth rate is translated into five different values as presented in Table 2 below.

The application of the method starts, once the data has been processed. Because PCA requires standardized input, all indications are transformed to $Z_{i,}$ Z-Score (see Eqn. (2)), and the processed data's mean value is 0, the standard deviation is 1.

$$Z_{i,m,n} = \frac{x - \overline{x}}{\sigma} \tag{2}$$

where, x is the indicator n of the stock m in year i, $\overline{x}$ is the average value of the indicator n in year i, and σ is the standard deviation of the indicator n in year i. The *n-dimensional* eigenvector matrix $\overrightarrow{\mathrm{U}}$ and the eigenvalue vector $\overrightarrow{\lambda}$ can be obtained. The next step is to sort the

Table 3
Six Stock Industry of (US Stock Market) used in this research for period 2015–2020.

Year	Public Utility	Commercial Industry	Industry	Property Sector	Comprehensive Industry	Finance	*Sum*
2015	305	110	1232	134	43	22	1846
2016	318	108	1203	135	37	19	1820
2017	378	123	1462	137	50	24	2174
2018	409	128	1635	146	55	21	2394
2019	469	140	1928	158	61	18	2774
2020	478	144	1992	159	56	18	2847

eigenvalues from largest to smallest, and select the eigenvalues with a cumulative contribution rate greater than 90% (i.e. $\Sigma_{i=1}^{j} \lambda i / \Sigma_{i=1}^{n} \lambda i \geq 90\%$). Sorting the eigenvectors according to the selected eigenvalues to obtain the Eigen vectors $\overrightarrow{u_1}, \overrightarrow{u_2}, \overrightarrow{u_3}, \ldots, \overrightarrow{u_J}$. To get the eigenvectors, multiply the original n standardized indicators variable data with the eigenvectors *j* data matrices for comprehensive principal component indicators $C_{i,m}$ (see Eqn. (3)). The weight $W_{i,}$ of each component is obtained by $\lambda_j / \Sigma_{i=1}^{n} \lambda_i$.

$$\begin{bmatrix} a_{11}, & a_{12}, \ldots, & a_{1n} \\ a_{21}, & a_{22}, \ldots, & a_{2n} \\ & \vdots & \\ a_{m1}, & a_{m2}, \ldots, & a_{mn} \end{bmatrix} \times \left[\overrightarrow{u}_1, \overrightarrow{u}_2, \overrightarrow{u}_3, \ldots, \overrightarrow{u}_j\right] = \overrightarrow{C}_{i.m} \tag{3}$$

Multiplying the *j* comprehensive principal component indicators data to obtained $C_{i,m}$, above by the weights $W_{i,}$ obtained above, and then add them together to get a total score. Minimum-maximum normalization (Eqn. (4)), a basic approach for pushing data into pre-defined bounds [C, D] [4,43], is used to increase the score to [0,100] in order to permit dynamic selection every year. Finally, the score is compared to a certain threshold, and the stocks with the highest scores are eliminated. It is important to remember that this phase must be regularly and dynamically updated each year based on the yearly financial reports to guarantee that a new scoring system can be formed in accordance with the stock market's development.

$$x = \frac{x - \min(x)}{\max(x) - \min(x)} (D - C) + C \tag{4}$$

whereas, C and D are the pre-defined bounds, max*(x)* is the maximum score and min*(x)* is the minimum score a stock received.

2.3. Algorithmic trading

Quantitative stock selection uses a year as the time dimension, whereas algorithmic trading must focus on the short period, such as monthly and daily. The machine learning method is used to forecast the stock price for the following month and determine the projected rate of return to represent the stock market's future features, therefore facilitating short-term trades. The closing price of the stock ***m*** in month ***h+1*** is predicted by the technical indicator of month ***h*** $(T_h,)$. 210 indicators are included and used to ensure that machine learning fully learns the laws of the stock market, including the closing price, opening price, highest price, lowest price, trading volume, turnover rate, and average price of the previous 20 days, as well as some derived indicators like MACD, KDJ, CCI, RSI, and others. These indicators are based on prior research and widely used stock market indicators [3,10,12,44]. A number of machine learning algorithms have been constructed and employed in previous research. Because the time period of prediction in this research is a month, which has been seldom employed in previous studies. As a result, the machine learning techniques employed in this study include SVM [2,45,46], ANN [4,5,29,47], and LSTM [36,26], which have been utilized in previous studies to discover the optimal answer. LSTM is a deep learning model that has seen increased application in the financial sphere in recent years due to its ability to forecast time series data accurately [26]. In this study, in order to maintain the consistency of the input, the input index $T_{h,m,t}$ is converted into $X_{h,m,t}$ through Minimum-maximum normalization, and its value ranges [0,1]. The model predicts the closing price at the end of the next month and compares it with the set threshold $r_{threshold}$ to determine the stocks purchased at the beginning of month *h+1*. However, depending just on forecasts to buy and hold is not comprehensive. Unlike prior research, this research will use the MACD and KDJ indicators as a guide for trading following a purchase. The MACD series is the difference between a "fast" (short period) and a "slow" (longer period) exponential moving average (EMA) of the price series. The MACD series can detect changes in a stock's trend by comparing EMAs from various periods. Table 3 present stock industry under consideration in the research.

The KDJ indicator takes into account not only the most recent closing price, but also the most recent high and low values. Which eliminates the flaw of just taking into account the closing price and neglecting true volatility. The return rates of MACD and KDJ paired with machine learning in the study to compared and identify which indicator better suited for the underline model. However, the model does not just rely on indicators, if actual return $\widehat{R}_{h+1,m}$ did not reach $r_{threshold}$ and will only be sold at the end of the month *h+1*. The factor loadings of the selected six stock industry are presented in Table 4 after transformation

Table 4
Factor Loadings of the Selected Six Stock Industry after transforming to normal variables.

Public Commercial Industry Property Comprehensive Finance Utility Industry	
Stock Returns	*Factor1 Factor 2 Uniqueness Factor1 Factor 2 Uniqueness Factor1 Factor 2 Uniqueness Factor1 Factor 2 Uniqueness*
r_{pu}	0.658 0.741 0.017 0.984 0.171 0.006 0.983 0.171 0.006–0.445 0.108 0.789 0.784 0.631 0.007
r_{CI}	0.841 0.534 0.006 0.202 0.967 0.048 0.201 0.945 0.046 0.693–0.359 0.382 0.831 0.568 0.006
r_I	0.806 0.548 0.048 0.350 0.935 0.007 0.351 0.944 0.007 0.992 0.111 0.008 0.728 0.654 0.049
r_{CI}	0.744 0.589 0.085 0.742 0.487 0.204 0.741 0.486 0.432 0.251–0.967 0.006 0.772 0.626 0.023
r_F	0.634 0.764 0.015 0.571 0.482 0.423 0.570 0.482 0.431–0.246 0.306 0.845 0.532 0.845 0.007

Note: $r_{pu}, r_{cI}, r_I, r_{cI}, r_F$ are stock returns.

Table 5
The number of selected stock based on PCA.

Year	2016	2017	2018	2019	2020	SUM
With EVA	44	36	34	32	28	174
Without EVA	40	32	28	36	24	160

3. Illustration of the model on US stock market

The proposed model is demonstrated on US A-share market in order to determine its stability and long-term capability, as well as to assess the benefits and drawbacks of machine learning.

3.1. Data and computational environment

A-shares have acquired increasing attention in the worldwide investing market as the U.S. economy has become stronger. As a result, this experiment is based on back-testing data from 2014 to 2020 from U.S. market indexes available at: https://www.cnbc.com/us-market-indexes Bloomberg available at: https://www.bloomberg.com and NYSE American available at: https://www.nyse.com/index. The rationale for this time frame is because annual financial reports. Every time the PCA investment portfolio is recreated, it is brought closer to real operation. Stock financial index data and EVA calculation data are sourced from NYSE American available at: https://www.nyse.com/index with equities with ST and missing data being omitted. Furthermore, all results reported in this research was carried out in Python. For Machine learning, we used the open-source computational Python library "NumPy", and for ANN and Deep learning we get helped from "TensorFlow" and "PyToch" Python Packages.

3.1.2. Parameters and evaluation of the model

This study's associated parameters are divided into two categories: thresholds and machine learning parameters. PCA's cumulative contribution rate is included in the form of 90% threshold $S_{threshold}$ (70) for screening scores, and the threshold $r_{threshold}$ (0.05) for algorithmic trading. This is how the latter is introduced. Sklearn3's default settings are SVR parameters. The ANN has four layers and 64 units, whereas the LSTM has three levels and 32 units. "ReLU" is their activation function, Mean Square Error (MSE) is their loss function (see Eqn. (5)), Mean Absolute Error (MAE) is their evaluation function (see Eqn. (6)), and Adam is their neural network optimizer [13].

$$\mathrm{MSE} = \frac{1}{m} \Sigma_{i=1}^{m} \left(y_i - \hat{y}_i\right)^2 \tag{5}$$

$$\mathrm{MAE} = \frac{1}{m} \Sigma_{i=1}^{m} \left|y_i - \hat{y}_i\right|^2 \tag{6}$$

3.2. Stock collection through PCA

As previously stated, multiple scoring models are developed for various businesses, and the scoring models will be updated on a regular basis. This study analyses the two investment portfolios with and without EVA indicators in order to determine the usefulness of the EVA indicators. The investment income of utilizing and not utilizing the EVA indicators will be compared at a later stage.

Table 5 displays the number of stocks chosen based on Principal Component Analysis (PCA) for each year and the number of stocks screened by EVA is higher than the number screened without EVA, indicating that the use of the EVA model allows for the mining of more valuable stocks. When a set criterion is used, less stock is chosen, while the total number of stocks in the market grows. The stronger the firm, the more investment it can acquire and the more profits it can earn, however a little and valuable firm may not be able to receive enough foreign investment and thus will not be able to receive it. The gap between the stronger and the weakest will intensify as the economy grows. It's institutional grouping, or institutional investors' herding behavior, which is described as a significant number of investment institutions trading a single stock in the same direction. The stocks are selected based on principle

Table 6
Performance evaluation matrix of the forecast models.

PCA - Without EVA		PCA - With EVA		
Model	MAE	MSE	MAE	MSE
LSTM	1.843	50.682	2.046	59.348
ANN	2.02	75.402	2.213	61.188
SVR	150.811	24208.673	153.174	24860.988

Table 7
Quantitative stock investing model returns.

Model	Indicators	Return without-EVA	Return With- EVA
ANN	MACD	−0.06247	0.27165
	KDJ	−0.05857	0.25988
LSTM	MACD	0.1744	0.15585
	KDJ	0.17878	0.163
SVR	MACD	−0.17029	−0.16808
	KDJ	−0.1308	−0.13476
SSI-300 Benchmarks		9.67%	9.67%

component analysis.

4. Results of machine learning models

In previous research, the time span of stock price forecasts was always in days, but this study tried to apply it to forecasts with a monthly dimension. The relevant parameters of the model have been described in section 3.1.2. Table 6 shows the errors of the three models' predictions of the closing prices one month later on the two portfolios. The four error values of the LSTM prediction model are 1.843, 50.682, 2.046, and 59.348. The LSTM forecast is more accurate over a longer time period, whereas the ANN forecast is less accurate but not significantly different from LSTM. This might be an issue with a model parameter and SVR's performance as a typical machine learning model is exceedingly low in this case, indicating that SVR is not suited as a stock prediction model over the time span of prediction. This inaccuracy will have a significant impact on the return rate of following transactions.

According to the model's architecture, after projecting monthly income in algorithmic trading next step is to execute daily trading using the MACD and KDJ indicators. Determine which stocks will have a 5% return at the end of the month, and buy them at the start of the month, using MACD and KDJ as stop-loss indicators, respectively, and clearing the position at the end of the month. As indicated in Table 7, convert the five-year back-testing result into an annualized rate of return. Because the purpose of stock investing is to make a profit, it's only reasonable to use the general market return as a comparison. The NYSE American "https://www.nyse.com/index" is used as the benchmark in this article.

The investment portfolio created with EVA has a better return rate, as shown in Table 7, and the best portfolio has a return rate of 27.165%, which is a very high return rate. Although LSTM-no EVA is greater than LSTM-EVA, the difference is not significant. This demonstrates how EVA, as a tool for evaluating an enterprise's internal worth, may assist investors in finding more valuable companies. The SVR's rate of return is negative regardless of whether the EVA factor is utilized, indicating that its prediction inaccuracy has had a significant impact on the model's rate of return and that it cannot be employed as a medium- and long-term prediction engine for stock market learning tools. Although the ANN model has the best rate of return, its prediction is not particularly stable, thus it should be used in conjunction with other machine learning approaches to improve it. LSTM may provide significant profits regardless of whether there is an EVA factor portfolio. This demonstrates that LSTM, as a deep learning representative, is well-suited to quantitative investment as a tool for predicting long-term stock market trends.

The indicators show that MACD and KDJ perform better than the others, implying that the difference between the two in actual operation is not very large. The annualized return of the SSI-300 index from 2014 to 2021 is 9.667%, which is much lower than the return rate (16.300%) of using LSTM-EVA when compared to the benchmark data. The comparison between each combination and the benchmark index is demonstrated in Fig. 2.

It is more natural to see that using EVA can effect in a more reliable rate of return, and that the high rate of return of LSTM-no EVA can be attributed to an involuntary alternative. In the previous seven years, the stock market in the United States has seen a bull market, a bear market, and a turbulent era, but employing the EVA factor model may generate a high and consistent income while avoiding the SVR model owing to severe forecast errors. This shows that the EVA and machine learning model described in this work is a useful, long-term model that takes into account bull markets, bear markets, and market volatility.

5. Conclusion and suggestions

In this research, a long-term quantitative stock investing model based on machine learning and EVA factors is suggested, which is one of the first attempts in quantitative investment. This model rates stocks in many industries dynamically every year, and it employs

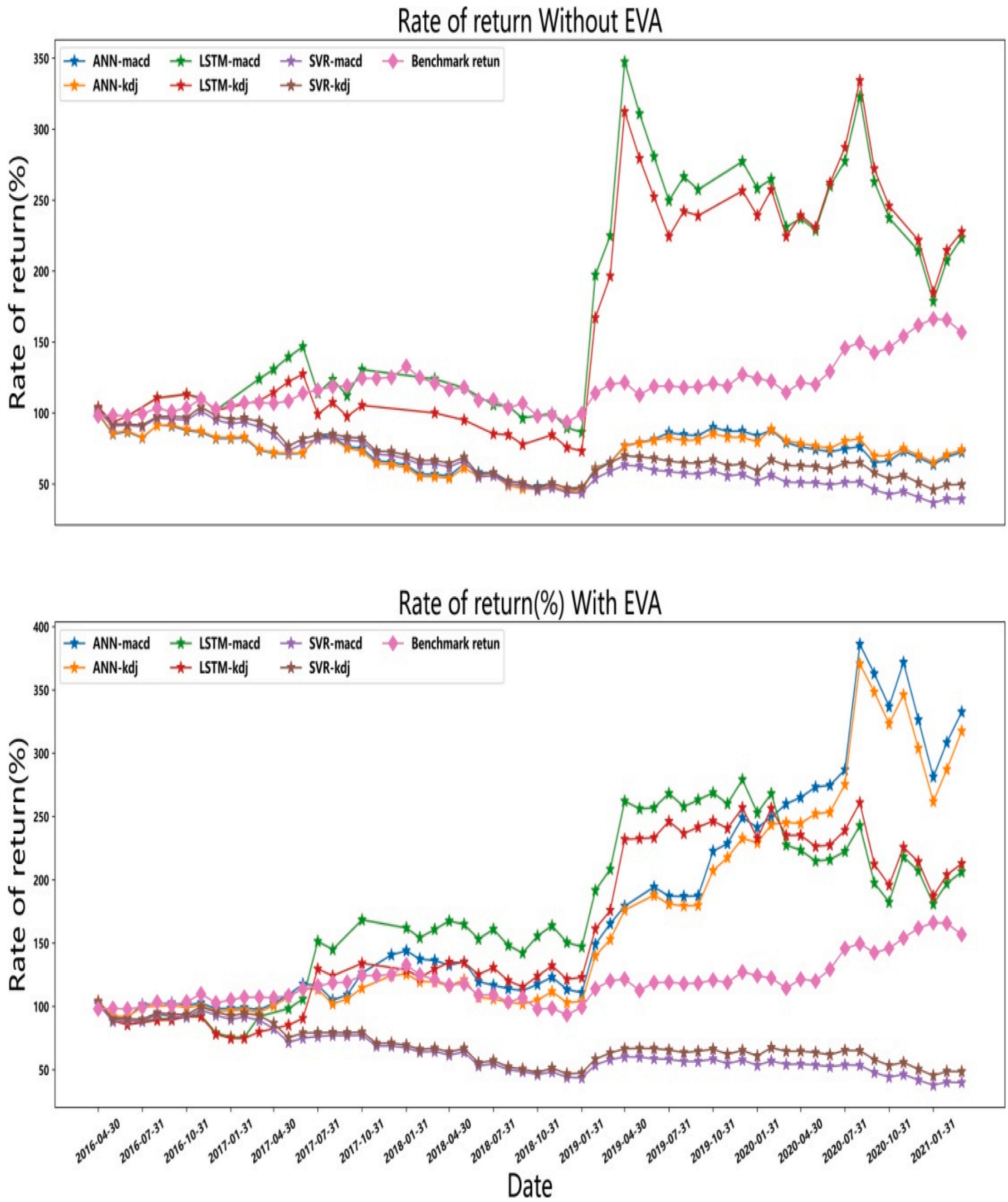

Fig. 2. Performance-testing of quantitative stock investment model of stocks returns.

a machine learning model to grasp the monthly rule, as well as technical indicators, to achieve long-term consistent high returns. The Matthew effect was progressively developed in the U.S stock market over the course of five years of back-testing, reminding U.S regulatory authorities of the need to manage the market and assist it perform its constructive function. EVA indicators may assist investors not only in obtaining a high and consistent return, but also in making sensible investments and promoting the stock market's proper capital allocation function.

The machine learning algorithm was employed in the long-term stock prediction model, the proposed ANN and deep learning model, should be given more attention, while the classic support vector regression (SVR) is no more efficient in presence of ANN and Deep learning approaches [41,28]. At the same time, investors should treat technical signs rationally rather than believing or abandoning them. The concepts of rational investment and value investment should be followed. The whole stock market will be more successful and steady development through machine learning, which is practical technology that helps people make logical decisions, via EVA, which is a true evaluation of the enterprise internal value of the component mining valuable stocks.

5.1. Future research

There are several aspects of this study, one may consider in future research for the purpose of improvement. First, the scoring model uses a limited number of criteria, and no judgment is made as to which elements are more important, necessitating more research at a later stage. Second, Machine learning models, particularly ANN, are not very stable. The accuracy of prediction will be influenced by the usage of set parameters. Auto ML technology might be applied at a later date to improve the situation. Finally, at the moment, the model is being applied for empirical research in the U.S market, but it is being explored for usage in the US market in the future.

Author contribution statement

Jiaqi Li, Xiaoyan Wang, Saleem Ahmad: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Xiaobing Huang and Yousaf Ali Khan: Performed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Data availability statement

Data will be made available on request.

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.

Abbreviations

MACD	Moving Average Convergence
PCA	Principal Component Analysis
EVA	Economic Value-Added
SI	Stochastic Indicators
ML:	Machine Learning
LSTM	Long-Short Term Memory
ROE	Return on Equity
ROA	Return on Assets
ANN	Artificial Neural Networks
SVM	Support Vector Machine
CAPM	Capital Asset Pricing Model
MSE	Mean Square Error
MAE	Mean Absolute Error
KDJ	Stochastic Indicator

References

[1] H. Moghaddam, M.H. Moghaddam, M. Esfandyari, Stock market index prediction using artificial neural network, Journal of Economics, Finance and Administrative Science 21 (21) (2016) 89–93.
[2] E. Sharifi Kazem, F.K. Hussain, M. Saberi, O.K. Hussain, Support vector regression with chaos-based firefly al- gorithm for stock market price forecasting, Applied Soft Com- puting Journal 13 (13) (2013) 947–958.
[3] P. Ratto, S. Merello, L. Oneto, Y. Ma, L. Malandri, E. Cambria, Ensemble of technical analysis and machine learning for market trend prediction, in: Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence, 2019.
[4] U. Levin, Stock selection via nonlinear multi-factor models, Adv. Neural Inf. Process. Syst. (8) (1996) 966–972.
[5] F. Huang, B.R. Chang, D.W. Cheng, C.H. Chang, Feature selection and parameter optimization of a fuzzy-based stock selection model using genetic algorithms, Int. J. Fuzzy Syst. (14) (2012) 65–75.
[6] G. Gounder, M. Venkateshwarlu, Shareholder Value Creation: an Empirical Analysis of Indian Banking Sector, Accounting and Finance Research, 2017.

[7] Huang, A hybrid stock selection model using genetic algorithms and support vector regression, Applied Soft Computing Journal (12) (2012) 807–818.
[8] H. White, Economic Prediction Using Neural Networks: the Case of IBM Daily Stock Returns, 1988, pp. 451–458.
[9] J. Huang, D.X. Yang, Y.T. Chuang, Application of wrapper approach and composite classifier to the stock trend prediction, in: Expert Systems with Applications, 2008.
[10] C. Stoean, W. Paja, R. Stoean, A. Sandita, Deep architectures for long-term stock price prediction with a heuristic-based strategy for trading simulations, PLoS One 14 (2019) 1–19.
[11] D. Chandwani, M. Singh Saluja, Stock direction forecasting techniques: an empirical study combining machine learning system with market indicators in the Indian context, Int. J. Comput. Appl. (2014).
[12] D.P. Kingma, J.L. Ba, Adam, A method for stochastic op- timization,3rd International Conference on Learning Representations, ICLR-Conference Track Proceedings 1– 15 (2015).
[13] E.F. Fama, K.R. French, Business conditions and expected returns on stocks and bonds, J. Financ. Econ. (25) (1989) 23–49.
[14] E. Fama, J.D. MacBeth, Risk, return, and equilibrium: empirical tests eugene F . Fama and james D. M ac- beth, Journal of Politcal Economy (1973) 607–636.
[15] F.M. Song, A two-factor ARCH model for deposit institution stock returns, J. Money Credit Bank. (26) (1994) 323.
[16] F. Yang, Z. Chen, J. Li, L. Tang, A novel hybrid stock selection method with stock prediction, Applied Soft Computing Journal 80 (2019) 820–831.
[17] G.S. Atsalakis, K.P. Valavanis, Surveying stock market forecasting techniques – Part II : soft computing methods, Expert Syst. Appl. 36 (2009) 5932–5941.
[18] H. Ahmed, Earnings and economic value added on the market share value: an empirical study on the islamic banks in Bangladesh, Global J. Manag. Bus. (2015).
[19] H. Markowitz, Portfolio selection, J. Finance 7 (1952) 77.
[20] J. Gomes, L. Kogan, L. Zhang, Equilibrium cross section of returns, J. Polit. Econ. 111 (2003) 693–732.
[21] J. Li, R. Zhang, Dynamic weighting multi factor stock selection strategy based on XG boost machine learning algorithm, Proceedings of 2018 IEEE International Conference of Safety Produce Informatization (2019), 868–872.
[22] V. Montano, T. Jombart, An Eigenvalue test for spatial principal component analysis, BMC Bioinf. (18) (2017) 1–7.
[23] W. Chen, Y. Cai, K. Lai, H. Xie, A topic-based sentiment analysis model to predict stock market price movement using Weibo mood, Web Intell. (14) (2016) 287–300.
[24] Y. Kara, M. Acar Boyacioglu, Ö.K. Baykan, Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines: the Sample of the Is- Tanbul Stock Exchange, Expert Systems with Applications, 2011.
[25] Y. Du, Application and analysis of forecasting stock price index based on combination of ARIMA model and BP neural network, in: Proceedings of the 30th U.S Control and Decision Conference, CCDC, IEEE, 2018, 2854–2857.
[26] T.C. Chiang, J. Li, L. Tan, Empirical investigation of herding behavior in U.S stock markets: evidence from quantile regression analysis, Global Finance J. (2010) 111–124.
[27] W.F. Sharpe, Capital asset prices: a theory of market equilibrium under conditions of risk, J. Finance 19 (1964) 425–442.
[28] D.P. Kingma, J.L. Badam, A method for stochastic op- timization,3rd International Conference on Learning Representations, ICLR-Conference Track Proceedings 1– 15 (2015).
[29] P. Gompers, J. Ishii, Corporate governance and equity prices, Q. J. Econ. 118 (2003) 107–155.
[30] O. Bustos, A. Pomares-Quimbaya, Stock market movement forecast: a Systematic review, Expert Syst. Appl. 156 (2020), 113464.
[31] Y. Fu, S. Cao, T. Pang, A sustainable quantitative stock selection strategy based on dynamic factor adjustment, Sustainability (12) (2020) 1–12.
[32] Y. Xu, C. Yang, S. Peng, Y. Nojima, A Hybrid Two-Stage Financial Stock Forecasting Algorithm Based on Clustering and Ensemble Learning, Applied Intelligence, 2020.
[33] K. Goel, S. Oswal, EVA-Based Financial Performance Measurement: an Evidential Study of Selected Emerging Country Companies, Corporate Ownership and Control, 2020.
[34] L. Tan, T.C. Chiang, J.R. Mason, E. Nelling, Herding behavior in U.S stock markets: an examination of A and B shares, Pac. Basin Finance J. (16) (2008) 61–77.
[35] M. Qiu, Y. Song, F. Akagi, Application of Artificial Neural Network for the Prediction of Stock Market Returns: the Case of the Japanese Stock Market, Chaos, Solitons and Fractals, Elsevier Ltd, 2016, pp. 1–7.
[36] S.G.K. Patro, K.K. sahu, Normalization: a preprocessing stage, Iarjset (2) (2015) 20–22.
[37] T. Suzuki, Y. Ohkura, Financial technical indicator based on chaotic bagging predictors for adaptive stock selection in Japanese and American markets, in: Physica A: Statistical Mechanics and its Applications, Elsevier B.V., 2016, pp. 50–66.
[38] J.M. Stern, G.B. Si, D.H. Chew, Eva: an integrated financial management system, Eur. Financ. Manag. (2) (1996) 223–245.
[39] M. Gowthul Alam, S. Baulkani, Local and global characteristics-based kernel hybridization to increase optimal sup- port vector machine performance for stock market prediction, in: Knowledge and Information Systems, Springer Lon- don, 2019, pp. 971–1000.
[40] S.A. Ross, The arbitrage theory of capital asset pricing (Work- ing Paper Version), J. Econ. Theor. 13 (1976) 341–360.
[41] X. Zhong, D. Enke, Forecasting daily stock market return using dimensionality reduction, in: Expert Systems with Applications, Elsevier Ltd, 2017, pp. 126–139.
[42] J. Lakonishok, A. Shleifer, R.W. Vishny, Contrarian investment, extrapolation, and risk, J. Finance 49 (1994) 1541–1578.
[43] S. Liu, C. Zhang, J. Ma, CNN-LSTM neural network model for quantitative strategy analysis in stock markets, Lecture Notes in Computer Science LNCS (2017) 198–206.
[44] K.J. Kim, Financial time series forecasting using support vector machines, Neurocomputing (55) (2003) 307–319.
[45] R. Zhang, Z. Lin, S. Chen, Z. Lin, X. Liang, Multi-factor stock selection model based on kernel support vector machine, J. Math. Res. 10 (2018) 9.
[46] X.Y. Fu, J.H. Du, Y.F. Guo, M.W. Liu, T. Dong, X.W. Duan, A Machine Learning Framework for Stock Selection, 2018 (arXiv).
[47] R. Peachavanish, Stock selection and trading based on cluster analysis of trend and momentum indicators, Lect. Notes Eng. Comput. Sci. 1 (2016) 317–326.