Machine learning-based approaches for financial market prediction: A comprehensive review

Bhaskar Nandi[1], Subrata Jana[2,*], Krishna Pada Das[3]

[1] *Department of Computer Science, Seacom Engineering College, Howrah 711302, West Bengal, India*

[2] *Jadavpur University, Kolkata 700032, West Bengal, India*

[3] *Department of Mathematics, Mahadevananda Mahavidyalaya, Barracpore 700120, Kolkata, West Bengal, India*

* Corresponding author: Subrata Jana, 87subratajana87@gmail.com

ABSTRACT: This research paper investigates the use of machine learning techniques in financial markets. The paper provides a comprehensive literature review of recent research on machine learning applications in finance, including stock price prediction, financial time series forecasting, and portfolio optimization. Various machine learning techniques, such as regression analysis, decision trees, support vector machines, and deep learning, are discussed in detail, with a focus on their strengths, weaknesses, and potential applications. The paper also highlights the challenges associated with machine learning in finance, such as data quality, model interpretability, and ethical considerations. Overall, the paper demonstrates that machine learning has significant potential in finance but calls for further research to address these challenges and fully explore its potential in financial markets.

KEYWORDS: financial markets; literature survey; research paper; machine learning; deep learning

1. Introduction

Financial markets are complex systems that involve the exchange of assets, such as stocks, bonds, and commodities, among buyers and sellers. The behavior of financial markets is influenced by a multitude of factors, such as economic indicators, political events, investor sentiment, and company news. The ability to predict and understand the behavior of financial markets is crucial for investors, traders, and policymakers alike.

In recent years, the use of machine learning (ML) in finance has gained significant attention due to its potential to extract insights from vast amounts of data and provide accurate predictions. ML techniques can be used to analyze financial data, identify patterns and trends, and make predictions based on historical data.

The use of ML in financial markets has numerous applications, including stock price prediction, fraud detection, risk management, and portfolio optimization. Additionally, ML has the potential to identify previously unknown market inefficiencies and opportunities, enabling investors to gain a competitive edge.

Despite the promise of ML in finance, the use of these techniques also poses significant challenges. These include issues related to data quality, overfitting, interpretability, and ethical considerations. Therefore, understanding the strengths and limitations of ML techniques in financial markets is crucial for ensuring their effective and responsible use.

This research paper aims to provide an overview of the current state of the art in using ML techniques in financial markets. The paper will explore the various applications of ML in finance, examine the challenges associated with these techniques, and provide recommendations for future research.

2. Literature survey

Strategies for improving an existing machine learning tool, C4.5, to handle concept drift and non-determinism in a time series domain were explored by Harries and Horn[1]. For the most part, human endeavors are plagued by the challenge of trying to foresee the future. Even though there are a variety of specialized time series projection methods available, they all have their drawbacks. In addition to being difficult to comprehend even for subject specialists, most methods are limited to modeling complete sequences rather than allowing users to derive predictive characteristics. In theory, symbolic machine learning could overcome these restrictions. The application of symbolic machine learning to a wide variety of difficult issues has proven highly effective. Unfortunately, there haven't been many efforts to explicitly apply symbolic machine learning to time series projection. As a consequence, current systems are unable to deal with evolving target ideas or clearly depict instances in a time-ordered fashion. Due to its temporal ordering, its target ideas' dynamic nature, and its high degree of non-determinism, financial projection is a difficult target domain. Many experts believe the financial markets to be uncertain, so the prospect of finding a method that is more reliable than random chance is appealing. The purpose of this research is to show that machine learning can be used to generate helpful financial forecasting techniques. The minimum usable success rate, in the eyes of subject specialists, for short-term financial prediction is 60%. Our findings suggest that, with the right set of tools, machine learning can achieve even better outcomes than this. We reduced the impact of both noise and idea shifts by compromising coverage for precision. Australian Gilt Securities Limited collaborated on and financed the described study. In this case, an Australian Postgraduate Award helped Michael Harries pursue his studies (industry).

Support vector machines (SVMs) were compared to a multi-layer perceptron trained with the back propagation (BP) algorithm in a study by Cao and Tay[2]. The parameters of normalized mean square error (NMSE), mean absolute error (MAE), directional symmetry (DS), correct up (CP) trend, and correct down (CD) trend all indicate that SVMs provide more accurate predictions than BP. The data collected is the monthly closing prices for the S&P 500. Due to the lack of a systematic method for selecting the SVM's free parameters, this exercise examines the generalization error with regard to the SVM's free parameters. This is demonstrated experimentally, where their effect on the answer is shown to be negligible. The benefits of using SVMs for finance time series forecasting are highlighted by the analysis of the experimental findings.

Kim[3] used support vector machines (SVMs), which are promising methods for the prediction of financial time series because they use a risk function consisting of the empirical error and a regularized term that is derived from the structural risk minimization principle. In order to forecast the stock market indicator, SVM is used in this research. Further, by contrasting SVM with back-propagation neural networks and case-based reasoning, this research looks at the viability of using SVM in financial predictions. It is clear from the experiments that SVM is a viable alternative to traditional methods of stock market forecasting.

In 2005, Huang et al.[4] submitted that the capacity control of the decision function, the use of kernel functions, and a sparse solution are what set support vector machines (SVMs) apart from other learning algorithms. In this study, we use SVM to anticipate the weekly direction of the NIKKEI 225 index's movement in order to explore the predictability of financial movement direction. We compare SVM's

predicting abilities to those of LDA, QDA, and Elman Backpropagation Neural Networks. The experimental outcomes favor SVM over competing categorization strategies. In addition, we suggest a merging model in which SVM is combined with the other categorization techniques. Out of all the available predicting techniques, the merging model is the most effective.

Predictions for the National Stock Exchange's S&P CNX NIFTY market index were analyzed by Kumar and Thenmozhi[5]. The NSX is one of the most rapidly expanding stock exchanges in developing Asian countries. Unique machine learning approaches, such as random forest and support vector machines (SVM), show promise in the forecasting of economic time series. Linear discriminant analysis, logit, artificial neural networks, random forests, and support vector machines are some of the categorization models used to make directional predictions that have been put to the test. In this research, empirical evidence indicates that the SVM is superior to the other classification methods for predicting the direction of stock market movement, while the random forest method is superior to the neural network, discriminant analysis, and logit model.

Stock and index price direction, accurately predicted by Ou and Wang[6] is critical for market dealers or investors to maximize profits. Data mining techniques have been successfully shown to generate high forecasting accuracy for stock price movements. Traders can no longer rely on a single method of predicting the future of the markets; instead, they must employ a variety of forecasting methods to obtain multiple indications and more information. Ten data mining methods are presented and used to forecast the performance of the Hang Seng indicator on the Hong Kong stock exchange. Methods such as Nave Bayes based on kernel estimation, Logit model, tree-based classification, neural network, Bayesian classification with Gaussian process, support vector machine (SVM), and least squares support vector machine are among those used. (LS-SVM). Based on the experimental findings, the SVM and LS-SVM produce the best predictive scores. In particular, SVM outperforms LS-SVM in terms of hit rate and error rate parameters for in-sample prediction, while LS-SVM outperforms SVM for out-of-sample predictions.

The use of BFO (Bacterial Foraging Optimization) and ABFO (Adaptive Bacterial Foraging Optimization) techniques was first introduced by Majhi et al.[7] to create an effective forecasting model for predicting different stock indices. These prediction models employ a straightforward linear combiner structure. As a result, ABFO and BFO are used to improve the linking weights of the models based on the adaptive linear combiner in order to produce the lowest possible mean square error (MSE). Comparisons are made between the findings obtained by these models and those obtained by the genetic algorithm (GA) and particle swarm optimization (PSO)-based models when evaluating their short- and long-term forecast performance with test data. In comparison to other evolutionary computing models like GA and PSO-based models, the new models are generally seen to be more computationally effective, prediction-wise more precise, and to demonstrate quicker convergence.

A method for predicting Indian stock market indices was presented in 2012 by Mohapatra et al.[8] who used a functional link artificial neural network (FLANN) based on differential evolution. The model predicts the S&P 500 and Dow Jones industrial average stock price indexes one day, one week, two weeks, and one month into the future using the back-propagation (BP) algorithm and the differential evolution (DE) algorithm, respectively. The experimental data uses market values from the Bombay market exchange (BSE), the national stock exchange of India (NSE), the India futures exchange (INFY), and a few other exchanges as input. DE always achieves better results than the BP algorithm. Measures of success include the mean absolute percentage error (MAPE) and the root mean square error (RMSE). The results show that DE has much lower MAPE and RMSE than the BP technique. Java-6 and NetBeans were used to carry out the computer analysis.

Standard & Poor's 500 (S&P 500) stock indicators (stock price and traded volume) were compared with the daily volume of tweets mentioning S&P 500 stocks at the market, sector, and company levels by Mao et al.[9] in August. In addition, we use Twitter data as external input in a linear regression with an exogenous input model to forecast stock market indicators. Our early data shows that there is a connection between the everyday volume of tweets and various measures of stock market activity. What's more, it seems that using Twitter to make stock market predictions is useful. In particular, we find that including Twitter data in the algorithm improves its ability to forecast whether the S&P 500 ending price will go up or down.

Using an ordinal data set that was transformed, Siew and Nordin[10] investigated the theory and application of regression methods for forecasting the direction of stock prices. Currency values and financial ratios are among the many different data categories found in the initial retransformed data source. Stock prices can be calculated using the information forms in monetary amounts and financial ratios. The transformed dataset consists entirely of ordinal data, which allows for a uniform method of rating stock price movements. Both procedures' results are analyzed and evaluated. They use WEKA (Waikato Environment for Knowledge Analysis), a collection of machine learning algorithms for data mining tasks and a machine learning program, to conduct regression analysis for the main design. Our study environment is the ups and downs of the financial market on the Bursa Malaysia exchange. The balance sheet, revenue statement, and cash flow statement from the company's most recent yearly report serve as the basis for this analysis. The stock market trading basic analysis methodology was used to create the factors included in the dataset. In order to generate these results, WEKA classifiers were used as methods. This study showed that the outcomes of regression techniques can be improved for the prediction of stock price trends by using a dataset in standardized ordinal data format.

Shen et al.[11] predicted that the stock market is a long-time attractive topic to researchers from different fields. Specifically, many studies have been done to forecast the direction of the stock market with the help of machine learning methods like support vector machines (SVM) and reinforcement learning. Using support vector machines (SVM), we suggest a novel prediction method that takes advantage of the temporal connection between international stock markets and different financial products to forecast the direction of stocks for the following day. The numerical findings show that for the NASDAQ, the S&P 500, and the DJIA, the precision of the predictions is 74.4%, 76%, and 77.6%, respectively. Identical algorithms are used with various regression algorithms to track the true growth of marketplaces. The suggested forecast algorithm is then evaluated in comparison to other standards using a rudimentary trading model.

A model for predicting stock market prices is suggested by Kazem et al.[12]. This model makes use of chaos mapping, the firefly algorithm, and support vector regression (SVR). There are three phases to the projection algorithm. In the first step, hidden dynamics in phase space are reconstructed using a delay coordinate embedding technique. Optimizing SVR hyperparameters is the focus of the second phase, which makes use of a chaos firefly algorithm. The third and final step involves using the improved SVR to make stock price predictions. The suggested method is important in three ways. While earlier research has used a genetic algorithm (GA) to improve SVR hyperparameters, this one uses chaos theory and the firefly algorithm instead. Second, the dynamics of phase space are reconstructed through a delay coordinate embedding technique. Third, because it uses systemic risk reduction, it is very good at making predictions. In order to demonstrate the efficacy of the suggested method, we applied it to three of the most difficult time series data sets available from the NASDAQ historical quotes database: Intel, National Bank shares, and Microsoft's daily closed (last) stock price. Mean squared error (MSE) and mean

absolute percent error (MAPE) scores show that the proposed model outperforms other methods of SVR, including genetic algorithm-based SVR (SVR-GA), chaotic genetic algorithm-based SVR (SVR-CGA), firefly-based SVR (SVR-FA), artificial neural networks (ANNs), and adaptive neuro-fuzzy inference systems (ANFIS). (MAPE).

In 2015, Patel et al.[13] tackled the challenge of forecasting the future direction of stock and stock price indices across 23 stock markets in India. The research evaluates and contrasts four forecast models using two different methods for inputting data: an artificial neural network (ANN), a support vector machine (SVM), a random forest, and a Naive-Bayes. Input data can be obtained in two ways: the first entails computing ten technical parameters from stock 26 transaction data (open, high, low, and close prices), while the second centers on depicting these 27 technical parameters as trend deterministic data. All 28 pairs of input methods are compared to the forecast models. Ten years of history data (2003–2012) are used to compare the performance of two companies (Reliance Industries and Infosys Ltd.) and two stock price benchmarks (CNX Nifty and S&P Bombay Stock Exchange (BSE) Sensex). For the 31st method of input data, where ten technical factors are depicted as continuous values, random forest appears to perform better than the other three prediction models, according to the findings of the experiments. Moreover, experimental 33 findings demonstrate that all prediction models benefit from representing these technical factors as 34 trend deterministic data, suggesting that this representation is superior.

Stock market index forecasting was the primary area of research for Patel et al.[14]. Two indices, namely the CNX Nifty and the S&P Bombay Stock Exchange (BSE) Sensex, from Indian stock markets are selected for experimental evaluation. The past data of these two metrics is used as the basis for the experiments, which span a period of 10 years. One-, ten-, and thirty-day forecasts are provided. In the first step of the paper's proposed two-stage fusion method, Support Vector Regression (SVR) is used. Second-stage fusion models combine SVR with predictions from an artificial neural network, a random forest, or another SVR model. These multi-stage models are compared to single-stage models (ANN, RF, and SVR) in terms of their forecast ability. A total of ten different technical metrics are used as inputs across all of the forecast algorithms.

Stock return forecasting with this innovative hybrid model was suggested by Rather et al.[15]. The suggested model incorporates two linear models—the autoregressive moving average model and the exponential smoothing model—and one non-linear model—a recurrent neural network. A novel classification algorithm is developed to produce training data for recurrent neural networks. Predictions made by a recurrent neural network are generally accurate, unlike those made by linear models. The suggested hybrid prediction model combines forecasts from the aforementioned three prediction-based models with the intention of further enhancing the precision of predictions. In this paper, we present a model for optimization that, when solved with genetic algorithms, produces optimum weights for the suggested model. The findings corroborate the strong predictive abilities of the recurrent neural network. To no one's surprise, the suggested hybrid prediction model beats recurrent neural networks in terms of forecast accuracy. When dealing with non-linear data whose patterns are challenging to capture with conventional models, the suggested model is widely regarded as a potential new direction to explore in the field of prediction-based models.

Normal distribution market microstructure (MM) models are helpful tools for modeling financial time series, but they do not account for skewness and heavy tails that may arise in a market, as opined by Xi et al.[16]. This article proposes a model of the heavy-tailed market substructure using the student-t distribution (MM-t) to address this issue. An effective Markov chain Monte Carlo (MCMC) technique is

devised for parameter estimation of the suggested model under the assumption of non-normality. The efficiency of the estimation method is confirmed by computer research. In an empirical analysis of several stock market benchmarks, the suggested model is compared to MM models using distributions like the normal distribution and a mixture of two normal distributions. For some financial time series, the MM-t model offers a better match than the MM models with other distributions, and empirical findings suggest that stock prices and returns have large tails. We also compare the MM-t model to another form of model, the stochastic volatility (SV-t) model with a student's t distribution, and find that the MM-t model provides the best match for the three indices.

To produce trading choices more efficiently, Dash and Dash[17] used a novel decision support system based on a computationally efficient functional link artificial neural network (CEFLANN) and a collection of rules. Here we formulate the issue of predicting stock trading decisions as a categorization problem, with the buy, hold, and sell signs as the three possible classes. By studying the nonlinear relationship between a small number of widely used technical indicators, the CEFLANN network used by the decision support system generates a series of continuous trading indications in the interval 0e1. In addition, the trend is followed, and trading decisions are made based on the trend using the trading signals produced by the system and a set of trading guidelines. The innovative aspect of the method is that it combines the technical analysis principles with the learning power of a CEFLANN neural network to predict when it will be most lucrative to buy or sell a company. The usefulness of the suggested method is evaluated by contrasting the results of the model with those obtained using other machine learning methods, such as the support vector machine (SVM), the naive Bayesian model (NBM), the K closest neighbor model (KNN), and the decision tree (DT) model.

With the aid of mathematical and statistical tools, Gerlein et al.[18] used technical and quantitative analysis in financial trading to assist buyers in determining when to place and cancel orders. While these more conventional methods have served their purpose, new methods have evolved from the field of artificial intelligence, such as machine learning and data mining, to analyze financial data. Simpler models that have proven their worth in fields other than financial trading are worth considering to ascertain their benefits and inherent limitations when used as trading analysis tools, even though the majority of financial engineering research has focused on more complex computational models like neural networks and support vector machines. Through a set of FOREX market trading simulations, this study examines the function of basic machine learning models in generating trading profits. It evaluates how well the models work and how certain model configurations generate reliable trading forecasts. We discuss the importance of attribute selection, periodic retraining, and training set size in generating positive cumulative returns for each machine learning model, and we show that simple algorithms, which have previously been excluded from financial forecasting for trading applications, can achieve comparable results to their more complex counterparts. The paper explains how the classification capabilities, which directly affect the final profitability, can be improved by using a combination of attributes in addition to technical indicators that have been used as inputs for the machine learning-based predictors, such as price-related features, seasonality features, and lagged values used in classical time series analysis.

If you're an investor or dealer, you need a reliable stock price prediction in order to make educated decisions, as said by Labiad et al.[19]. As a result of their nonlinearity and nonstationarity, however, values exhibit a complicated pattern of behavior. In this article, we apply three machine learning methods—Random Forest (RF), gradient boosted trees (GBT), and support vector machines (SVM)—to forecast minute-by-minute changes in the Moroccan stock market. (SVM). To enhance prediction accuracy and

decrease training time, a number of technical markers were used as input variables, and a feature selection and sample selection phase were carried out. For this exercise, we use the intraday values (tick-by-tick data) of Maroc Telecom (IAM) shares over the course of eight years to compare the results of various models. Our experiments have demonstrated that RF and GBT are better than SVM on our dataset. Additionally, RF and GBT are well suited for short-term prediction due to their minimal computational complexity and decreased training time.

Market performance of Karachi stock exchange (KSE) on day closing by using various machine learning methods and it was predicted by Usmani et al.[20]. The market is predicted to be either positive or negative based on a variety of characteristics used in the prediction algorithm. In the model, we use variables such as the price of oil, the price of gold and silver, the interest rate, the FEX rate, the news, and a social media stream. Additionally, traditional statistical methods such as the simple moving average (SMA) and the autoregressive integrated moving average (ARIMA) are utilized. Single-layer perceptron (SLP), multi-layer perceptron (MLP), radial basis function (RBF), and support vector machine (SVM) are some of the machine learning methods examined here. Furthermore, each of these characteristics is examined independently. For the greatest results, use the MLP algorithm instead of the other methods. The correlation between gasoline prices and market success was the strongest. The outcomes indicate that the KSE-100 index's future success can be forecast using machine learning methods.

Financial time series has been one of the most difficult issues in financial market research for quite some time, and it is forecasted by Tsantekidis et al.[21]. Investors typically use statistical models to help them determine when it is optimal to join and leave the markets (or even simple qualitative methods). However, the used models' predicting precision is badly limited by the noisy and stochastic character of markets. In order to overcome some of the problems with the aforementioned methods, new machine learning techniques have been developed since the advent of computerized trading and the availability of large quantities of data. In this paper, we suggest a recurrent neural network-based deep learning technique for analyzing limit order book data at high frequency to forecast future price changes. A massive dataset of limit order book occurrences is used to assess the effectiveness of the suggested technique.

Chen and Hao[22] looked into the future performance of stock market benchmarks, a topic of interest and importance in the fields of finance and applications because it offers the possibility of higher returns with reduced risk through the use of prudent currency exchange policies. Many approaches have been attempted with varying degrees of success to achieve precise prediction, but it is the machine learning approaches that have attracted focus and development. In order to accurately forecast stock market benchmarks, we suggest a simple hybridized framework that combines the feature weighted support vector machine and the feature-weighted K-nearest neighbor. To begin, we develop a comprehensive theory of feature-weighted SVM for data categorization by establishing a hierarchy of feature weights based on their significance to the classification. Next, we compute the information gain for each characteristic to establish its relative significance and obtain the weights. Finally, we compute k-weighted nearest neighbors from the past data and use them as features in feature-weighted K-nearest neighbor to forecast future stock market values. In order to evaluate the efficacy of our developed model, we show experimental findings on two prominent Chinese stock market indices, namely the Shanghai and Shenzhen stock exchange indices. Our suggested model improves upon prior methods for forecasting the short-, intermediate-, and long-term performance of the Shanghai stock exchange composite index and

the Shenzhen stock exchange component index. The suggested algorithm is flexible enough to be applied to the forecasting of various stock market benchmarks.

Machine learning is a subfield of AI with the capacity to predict the future based on past experience, and it was used by Hitam and Ismail[23]. Machine learning techniques like neural networks (NN), support vector machines (SVM), and deep learning are among the methods suggested to build models. In this article, we evaluate the accuracy of various machine learning methods for predicting the price of digital currencies. In this article, we will be focusing on predicting time series data. Prior studies have shown that SVM gives a result that is almost or near the real result while also improving the accuracy of the result itself, so it is clear that SVM has a number of advantages over other models in forecasting. Recent studies have shown, however, that the general status and accuracy rate of forecasting need to be improved in order to be useful. This is because of the limited scope of the samples used and the manipulation of the data by insufficient proof and expert analysts. This necessitates an extensive study into how close actual prices come to the predicted ones.

The fixed income market is crucial to the economy, as opined by Martin et al.[24]. Corporate problems are related to sovereign issues, which are affected by central bank policy. Bonds are less accessible and less transparent than stocks, so there is less information available to the public about them. We show how machine learning models can be used to predict interest rates on U.S. Treasury securities of different terms and the clean values of business securities.

According to Reddy[25] stock trading is one of the most crucial operations in the realm of business. Speculating on the future price of a company or other financial asset traded on a stock exchange is known as stock market prediction. In this article, we explore how Machine Learning can be used to forecast market performance. Most stock traders use basic and fundamental analysis, or time series analysis, when forecasting stock prices. In order to make accurate stock market forecasts using machine learning, Python is the tool of choice. In this article, we advocate for a machine learning (ML) strategy that can be taught using publicly accessible stock market data and can then apply this knowledge to make reliable forecasts. This research employs daily and up-to-the-minute stock values from three marketplaces and two distinct market capitalization sizes to make stock price predictions using a machine learning method called support vector machines (SVM).

According to Ren et al.[26], the stock market is heavily influenced by trader opinion. In addition to stock market statistics, user-generated textual material on the Internet is a valuable source for reflecting investor behavior and predicting stock values. In this work, we combine mood analysis with a machine learning technique using a support vector machine. We also account for the weekday impact to create more credible and grounded mood indices. Adding mood factors can improve direction prediction accuracy for the SSE 50 Index by as much as 89.93%, with an additional 18.6 percentage points. Further, our approach aids buyers in making better choices. Furthermore, these results suggest that sentiment is one of the main indicators of the stock market because it likely includes valuable information about the intrinsic value of assets.

Long short-term memory (LSTM) neural network models and support vector machine (SVM) regression models were compared in research given by Lakshminarayanan and McCrae[27]. In total, eight different models went into the structure developed for this research. Here, we use LSTM to construct four models and SVM to construct four additional models. This study makes heavy use of two large databases. Two such examples are provided, one using only the normal stock price data from the Dow Jones index (DJI), and the other using that data as well as the external input factors of crude oil and gold

rates. A top model for our raw data is revealed by this comparison research. Root-mean-squared error, mean-squared error, mean-absolute error, mean-absolute percentage error, and R-squared are all measures of model success. We review the models' methods and outputs and offer suggestions for improving upon this work.

Big data analytics were used by Modi et al.[28] across many different fields to make precise predictions and conduct thorough analyses of massive datasets. They make it possible to unearth valuable insights that would otherwise be buried in mountains of data. In this article, we present a method for analyzing the stock market, allowing investors to gain insight into the market's volatility while also gaining a better understanding of how to benefit from it. To begin, we present a literature review of prior efforts in this field. We then detail our technique, which includes data gathering and machine learning algorithms.

Mohan et al.[29] have long been interested in attempts to predict stock market values. The high volatility of stock values makes them difficult to forecast, as they are sensitive to a wide range of political and economic factors, as well as shifts in leadership, investor mood, and other factors. Both past data and written information alone have proven to be inadequate for accurate stock price forecasting. Existing mood analysis research has discovered a robust relationship between stock price changes and the arrival of new stories. Support vector machines, naive Bayes regression, and deep learning are just a few of the many methods that have been tried in numerous mood analysis studies at varying levels. The quality of results produced by deep learning algorithms is proportional to the quantity of available training data. Previous studies have gathered and examined too little textual data, which has led to inaccurate forecasts. In this paper, we show how to use deep learning models for better stock price forecasting by collecting a large quantity of time series data and analyzing it alongside relevant news stories. Our dataset contains five years' worth of daily S&P500 stock values and over 265,000 stories from the financial press that mention these businesses. Due to the massive scale of the dataset, we find cloud computing to be an indispensable tool for training forecast models and running inferences on a stock in real time.

Reddy[30] says the stock market is volatile, so understanding how it will likely behave in the future will help investors plan their investments more effectively. An accurate prediction is a surefire way to increase your earnings significantly. Time series analysis-based modeling has been used to improve the forecast accuracy of numerous models suggested in the economics and finance literature. The primary purpose of this article is to use stochastic time series ARIMA modeling to test for stationarity in time series data and to forecast the direction of change in a stock market index. Taking into account the lowest values of AIC, BIC, RMSE, MAE, MAPE, standard error of regression, and adjusted R2, the best-suited ARIMA (0,1,0) model was selected for predicting the values of time series, viz. BSE_CLOSE and NSE_CLOSE. Based on the weekly data from 6 January 2014 through 31 December 2017, the best-fitting model was used to make forecasts for the time spanning 1–7 January 2018 (3 anticipated values). (187 observed values). The study's findings corroborated the use of the ARIMA model for short-term time series forecasting, which should help the investing community make more informed, lucrative choices.

According to research by Zhong and Enke[31], big data analytic methods involving machine learning algorithms are assuming a more pivotal position in many practical domains, including trading in the stock market. However, there has been a dearth of research into the topic of daily stock market return predictions, even when advanced machine learning methods like deep neural networks (DNNs) are used. DNNs use a wide range of deep learning methods that differ in how well they perform based on the specifics of the data representation format, network structure, activation function, and other model factors. This paper introduces a big data analytics method based on 60 financial and economic features to forecast the daily return path of the SPDR S&P 500 ETF (ticker symbol: SPY). To forecast the daily

trajectory of upcoming stock market index returns, we apply DNNs and conventional artificial neural networks (ANNs) to the full preprocessed but untransformed dataset, as well as to two datasets transformed via principal component analysis (PCA). With overfitting mitigated, a trend in the DNNs' categorization accuracy is observed and illustrated as the number of hidden layers rises from 12 to 1000. In addition, a battery of hypothesis testing procedures is applied to the classification, and the results of the simulations demonstrate that the DNNs trained on the two PCA-represented datasets provide significantly higher classification accuracy than the DNNs trained on the entire untransformed dataset and several other hybrid machine learning algorithms. Furthermore, the DNN classification process led by PCA-represented data outperforms the others evaluated, including a comparison against two conventional benchmarks.

Anticipated returns in the stock market are typically framed as a forecasting issue where prices are anticipated by Basak et al.[32]. International financial markets' inherent instability makes forecasting difficult. As a result, many of the difficulties associated with attempting to foresee future stock market movements are mitigated by the use of forecasting and diffusion modeling. Invest with less worry if you can reduce predicting inaccuracy. The present study frames the issue as a direction-predicting exercise with positive and negative outcomes. To test whether stock values will rise or fall relative to the price prevalent a few days ago, we designed an experimental approach for the categorization issue. Random forests and gradient-boosted decision trees (with XGBoost) are two methods that utilize groups of decision trees to enable this link. We evaluate our method and detail the increases in accuracy over previous forecasts for a number of different businesses. To accurately forecast the direction of stock prices over the medium to long term, the current study introduces a new approach to selecting technical indicators and using them as features.

Recently, palmprint recognition has gained popularity due to its high identification accuracy, cheap cost of hardware, and simplicity of use with low-quality images, as seen by Michele et al.[33]. Numerous image-extracted palmprint characteristics have been used for identification purposes. One disadvantage of conventional image-based biometric identification systems is that a lot of work has gone into designing and getting a pertinent collection of effective handcrafted features. To address this issue, this article investigates the feasibility of applying Mobile Net V2 deep convolutional neural networks to palmprint identification by fine-tuning a previously trained Mobile Net network. We also investigate dropout support vector machines (SVM) and their functionality by training them on the same deep features as the comparable pretrained networks. The Hong Kong Polytechnic University of Science and Technology provides the information used in the studies. There are 6000 128 × 128 grayscale pictures of 500 distinct hands included in the collection. State-of-the-art performance on the datasets is shown to be achieved by the suggested methods. The second method, using Mobile Net V2-based features and a support vector machine (SVM) classifier, outperforms the best previously published findings with an average testing and confirmation accuracy rate of 100%.

Stock price modeling and projection, as done by Long et al.[34], have been difficult goals for academics and investors due to the noisy and non-stationary features of samples. Feature learning is a job that has become more efficiently handled by a specially built network thanks to the development of deep learning. For the purpose of feature extraction on financial time series samples and price movement prediction assignment, this article proposes a new end-to-end model called a multi-filter neural network (MFNN). The multi-filter's structure is built using a combination of convolutional and recurrent neurons to acquire data from multiple feature areas and market perspectives. We use the CSI 300 indicator of Chinese stocks to put our MFNN through its paces in severe market prediction and signal-based trading

simulation jobs. The experimental findings demonstrate that in terms of precision, profitability, and stability, our network is superior to the conventional machine learning models, the statistical models, and the single-structure (convolutional, recurrent, and LSTM) networks.

Machine learning algorithms designed to achieve specific goals, as described by Berislav and Hrvoje[35] have quickly gained acceptance as useful resources for analyzing financial data and making stock price predictions. The purpose of this article is to use machine learning algorithms (linear regression, gaussian processes, SMOreg, and neural network multilayer perceptron) on historical data from February 1, 2010, to January 31, 2020, to make predictions for five major stock market indices (DAX, Dow Jones, NASDAQ, Nikkei 225, and S&P 500). We used a variety of past-based period lengths and forecast ranges to arrive at our predictions. Error measures were used to gauge the effectiveness of machine learning techniques. Analysis shows that machine learning systems performed exceptionally well as forecasters. For lower basis period lengths and forecast ranges, all systems' accuracy improved. It's possible that the analysis's findings will aid shareholders in settling on a workable investment plan. However, predicting stock prices is still one of the most difficult problems in business.

The value of various machine learning techniques for predicting time series on financial markets is seen by Ghasemzadeha et al.[36]. One major challenge is that economic administrators and the scholarly community are still hoping for more precise forecasting algorithms. When this is done, the accuracy of the forecasts improves, leading to greater profits and streamlined operations. While introducing the best features, this article will demonstrate how a financial time series of technical factors available on the Tehran stock market can be used to produce useful outcomes. The proposed strategy makes use of machine learning methods based on regression, with an emphasis on choosing the most salient characteristics to discover the optimal technical variables of the inputs. These processes were applied to Python-based machine learning tools. This paper's collection was comprised of stock information for two businesses listed on the Tehran Stock Exchange, covering the years 2008 through 2018. The experimental findings demonstrate that the top methods successfully determined optimal parameters for the algorithms based on the chosen technical characteristics. Using those numbers allows for market data predictions with a minimal margin of error.

However, stock markets are driven by volatile factors like microblogs and news, making it hard to predict stock market indexes based on merely historical data, as Khan et al.[37] have predicted. Given the extreme volatility of the stock market, it is essential to accurately evaluate the impact of exterior variables on stock forecasting. Since data from social media and financial news can influence investment behavior, it can be used by machine learning systems to forecast stock market movements. In this article, we apply algorithms to data from social media and financial news outlets to learn how this information affects the reliability of stock market forecasts over a ten-day period. The data sets undergo feature selection and spam tweet reduction to boost the efficiency and quality of forecasts. In addition, we conduct tests to identify unpredictable stock markets and those that are particularly susceptible to the effects of financial and social media news. To find a reliable classifier, we examine the outcomes of multiple methods and compare them. Finally, ensembles of classifiers and deep learning are used to improve forecast accuracy. Predictions made using social media have an accuracy of 80.53%, while those made using financial news have an accuracy of 75.16%, according to our experiments. We also demonstrate that social media has a larger impact on IBM and Microsoft shares in New York, while financial news has a greater impact on London's stock market. The aggregate of random forest classifiers is determined to be the most accurate (at 83.22 percent).

It has been predicted by Kilimci and Duvar[38] that the directions in which stocks, exchange rates, and stock markets will move are important and a busy study field for investors, analysts, and academics. Based on an examination of nine high-volume banking equities listed on the Istanbul stock exchange (BIST 100), this article proposes a method for doing direction prediction for the BIST 100 using word embeddings and deep learning. To the best of our knowledge, no previous work has combined deep learning techniques with word embedding methods to predict the direction of Turkish stocks and markets using either Turkish news articles or user comments from social media and other platforms. A variety of deep learning methods, including LSTM networks, RNNs, and CNNs, as well as word embedding models Word2Vec, GloVe, and FastText, are compared and contrasted to achieve this goal. Four Turkish news sources are compiled to illustrate the usefulness of the suggested paradigm. Company-related stories from the public disclosure platform (KAP), Bigpara's textual technical analysis of each company, and user feedback from Twitter and Mynet Finans are collected. The experimental findings show that the direction of BIST 100 can be predicted with high accuracy using a mix of deep learning techniques and word embedding methods.

Reviewing the literature, Obthong et al.[39] say that when investing in the stock market, quick access to reliable data is essential for making sound buying choices. As a result of the sheer volume of stocks exchanged on a stock market, there are a plethora of considerations that go into making any given investment. Furthermore, stock market behavior is unpredictable and difficult to forecast. Predicting the future value of a commodity is crucial but difficult for these reasons. This motivates studies aimed at identifying the most efficient forecast model, one that can make the most precise projection with the least amount of error. In order to better forecast stock prices, this article reviews research on the use of machine learning techniques and algorithms.

When constructing a portfolio, Chen et al.[40] found that the future performance of equity markets was the most important factor. Opportunities to use prediction theory in portfolio selection have increased significantly with recent advances in machine learning. Many studies, however, demonstrate that using just one prediction model is not enough to make highly precise forecasts and generate substantial profits. In this article, we create an innovative portfolio-building strategy by combining the machine learning model for stock forecasting with the mean-variance (MV) model for portfolio selection to create a hybrid model. The two main steps of this approach are market forecasting and portfolio management. In the first step, we suggest a composite model for stock price forecasting that combines extreme gradient boosting (XGBoost) and an enhanced firefly algorithm (IFA). To fine-tune XGBoost's settings, the IFA was created. Stocks with greater yield potential are chosen in the second phase, and the MV model is used to pick the portfolio. When applied to data from the Shanghai stock exchange, the findings show that the suggested approach outperforms both the status quo (methods that don't involve stock prediction) and industry standards in terms of return on investment and risk.

Janková[41] set out to conduct a comprehensive literature review on the topic of AI in the financial markets, focusing on its primary study areas, current growth trends, and major publications. This document makes use of the bibliographic database viewer, VOSViewer. Our results, based on an evaluation of 353 papers and comments culled from the Web of Science database, are as follows: Financial time series can be predicted using artificial intelligence tools like neural networks and fuzzy logic, and decision models can be developed using the results; the most frequently cited authors in this area are Markowitz and Lebaron. Much of the foundational study of artificial intelligence can be traced back to the publication of expert systems with applications. Using powerful bibliometric techniques, we offer in-depth analysis and a new perspective on the field of study, empowering researchers of all levels—

but particularly those just starting out—with the tools they need to make informed decisions about their work. Focusing on mixed models, which are becoming increasingly prevalent in the field, is advised for the study's forecast of specific financial market segments.

Predicting stock market patterns, as Prasad and Seetharaman[42] did, is challenging but rewarding work. With the advent of the GPU and the TPU, experts and academics are increasingly turning to cutting-edge methods like machine learning to foresee movements in the stock market. Several stock-trend prediction systems have been created in recent years. Reviewing machine learning research papers and analyzing the significance of their results in the context of how stock price patterns produce trading cues is becoming increasingly important for assisting investors interested in making short-term investments in the stock market. To accomplish this, the authors of this article reviewed over fifty scholarly works that explored different machine learning algorithms for a wide range of input variables. They discovered that while the performance of models as measured by root-mean-square error (RMSE) for regression and accuracy score for classification models varied greatly, the long short-term memory (LSTM) model displayed higher accuracy than the rest of the machine and deep learning models studied. When revenue and Sharpe ratio were used to evaluate algorithm performance, reinforcement learning algorithms came out on top. In most cases, machine learning can be more profitable for dealers than basic analysis. Although technical analysis is simple to execute, the resulting reward may evaporate too quickly, making making a profit with it seem almost impossible. Traders and buyers must, therefore, familiarize themselves with machine, deep, and reinforcement learning algorithms. These results are founded on the synthesis of the literature reviewed, which is presented in the following part.

To evaluate the accuracy of various classifications, Subasi et al.[43] provided a side-by-side analysis of their predictions. Not only that, but the contrast is made with precision in mind. Each machine learning technique is put to the test on the NASDAQ, NYSE, Nikkei, and FTSE 100. Additionally, normal and compromised data collection is used to evaluate several machine learning methods.

Anomaly detection using machine learning is an extensively researched subject across many different application areas, as discovered by Tiwari et al.[44]. Due to the scarcity of labeled data and the fact that aberrant behaviors are frequently contextual and distributed across a series of anomalous occurrences, the identification issue for market monitoring remains difficult. This study offers an in-depth analysis of the most recent developments in machine learning techniques, with a focus on their applications in the field of financial market monitoring. Here, we talk about the problems and solutions that have arisen in this area of study, which has found most of its practical use in adjacent fields. In this paper, we present a case study of a machine learning-based monitoring system design for a physical power trading market and explore how the input data impacts the techniques' ability to identify abnormal market behaviors. Our research shows that ensemble algorithms based on regression trees can accurately forecast values one day into the future, demonstrating their ability to identify anomalous price movements.

According to research by Beukel[45], stock success, growth, and danger all play significant roles in the stock market and worldwide economy. Machine learning shows promise as a method for achieving this. The danger and security of an investment in equity are measured by its price. People in the financial market are increasingly on the lookout for ways to increase their profits as a result of the widespread adoption of cryptocurrencies and the improvement in computing power. In this paper, I define machine learning, discuss the factors that go into establishing stock values, and show how the application of machine learning to the financial market has led to more precise valuations. Researchers' results are summarized, and a conclusion is made.

According to Chen[46], the volatility and susceptibility to external variables that profoundly influence investor mood make stock market research and forecasting a difficult issue for finance specialists. The most recent development in stock market prediction technologies is machine learning, which generates predictions based on the values of current stock market indices by training on their previous values. The findings, however, appear to be volatile and unreliable, and the prediction techniques and algorithms are still in their early stages of development. The financial markets are highly sensitive to how quickly and accurately forecasts can be made, so it's crucial that the models be refined and the range of machine learning techniques broadened. Through a qualitative summary of the results obtained from various existing sources and experiments, this paper examines the relative merits of four different models for stock market prediction: the support vector machine (SVM), the convolutional neural network (CNN), the regression-based model, and the long short-term memory (LSTM). This paper's findings suggest that SVM and a mix of CNN and LSTM are effective at forecasting stock prices. The stock market is the marketplace where investors purchase and sell shares of publicly traded businesses. The goal of every buyer and vendor in the stock market is to make as much money as possible while minimizing their losses. Stock price forecasting can benefit from the application of cutting-edge technologies like AI. Extreme volatility, non-linearity, and changes in both the internal and exterior environment define the stock market. Using AI methods to identify this nonlinearity allows for significantly enhanced prediction accuracy. Since the rapid growth of technology and widespread adoption of personal computers in the 1990s, the study of how Artificial Intelligence (AI) can be applied to financial finance has garnered a great deal of academic interest. The issue of predicting stock prices has been the subject of numerous solutions since then. Based on a selection of 2326 papers drawn from the Scopus website between 1995 and 2019, this paper provides a comprehensive analysis of the literature on the application of Artificial Intelligence to stock market trades. This capstone project utilized artificial intelligence (AI) tools for stock market analysis and prediction, as well as the establishment of complicated nonlinear connections between the input data and the output data. Portfolio optimization, AI-based stock market prediction, financial mood analysis, and hybrid methods combining two or more of these techniques were identified as distinct groups within these articles. Each subfield's study history from its inception to the present day is detailed. In addition, a survey of the literature suggests that this field of study is receiving growing attention and that previous studies have laid the groundwork for more recent ones to build upon.

The equity market is extremely unstable, as Jishtu et al.[47] pointed out. However, there are a number of methods and perspectives one can employ to study this shifting paradigm and get ready for it as technology advances. In this article, we will discuss a variety of techniques for detecting market patterns with relative ease. The proposed approach is all-encompassing due to its incorporation of pre-processing the stock market dataset, various feature engineering techniques, and a bespoke deep learning-based system for predicting stock market price trends. The forecast technique with the lowest error rate is the one most often recommended. For this research, we used three separate algorithms to analyze the tone of press coverage of the company and its shares. The classification's findings have provided investors with more data for making educated wagering choices and a sharper understanding of the market's erratic swings for both long- and short-term investors.

According to research by Soni and Srivastava[48], the social network database grows exponentially as a result of the constant flow of new feedback, remarks, and messages. In order to understand how customers feel about a business or its offerings, it is now essential to sift through mountains of data. While English remains the most common language for web reviews, technological advancements and increased literacy have led to an increase in Hindi-language content. In addition, understanding how people feel

about a product is crucial to Indian language mood analysis, and we give credence to everyone's opinions. We used the Hindi language repository for general news items from multiple news sources to improve the precision of our classification. Naive Bayes, in addition to other machine learning categorization methods, such as random forest, Support Vector Machines, and Logistic Regression, were studied for their ability to accurately categorize texts.

3. Research methodology

The research methodology for machine learning in the finance sector involves several steps to ensure that the research is conducted in a systematic and rigorous manner. The following are the general steps that are typically followed in research on machine learning in finance:

3.1. Research question and hypothesis

The first step in any research project is to identify a research question and develop a hypothesis that can be tested using machine learning techniques. The research question should be relevant to the finance sector and address a current problem or challenge.

3.2. Data collection

The next step is to collect the relevant data that will be used to train and test the machine learning models. The data can be obtained from various sources, including financial statements, stock prices, news articles, and social media.

3.3. Data pre-processing

Once the data is collected, it needs to be pre-processed to remove any irrelevant or duplicate data and to convert the data into a format that can be used by the machine learning algorithms. This step involves cleaning, transforming, and normalizing the data.

3.4. Feature selection

In this step, relevant features are selected from the pre-processed data that will be used to train the machine learning models. Feature selection is an important step, as it reduces the complexity of the model and improves its performance.

3.5. Model selection

Once the features are selected, the next step is to select the appropriate machine learning model that will be used to analyze the data. The model selection depends on the nature of the problem and the type of data.

3.6. Model training and testing

In this step, the selected model is trained using the pre-processed data, and its performance is evaluated using a testing dataset. The performance of the model is evaluated using various metrics such as accuracy, precision, recall, and F1-score.

3.7. Result interpretation

The final step is to interpret the results of the machine learning model and draw conclusions based on the research question and hypothesis. The results are then discussed in the context of existing literature and potential future research directions.

Overall, the research methodology for machine learning in finance involves a rigorous process of data collection, pre-processing, feature selection, model selection, training, testing, and result interpretation. These steps ensure that the research is conducted in a systematic and transparent manner and that the results are reliable and valid.

4. Conclusion

In conclusion, the use of machine learning in financial markets has shown significant promise in recent years, offering new opportunities for investors, traders, and policymakers to better understand market behavior and make more informed decisions. ML techniques can be used to extract insights from vast amounts of financial data, identify patterns and trends, and make predictions based on historical data.

However, the application of ML in finance also poses significant challenges, including issues related to data quality, overfitting, interpretability, and ethical considerations. These challenges require careful consideration and attention to ensure that the benefits of using ML in finance are realized while minimizing potential risks and harms.

Looking forward, further research is needed to address these challenges and explore the full potential of ML in financial markets. Future research could investigate the development of more robust ML models that can address issues related to data quality and overfitting. Additionally, research could focus on developing more interpretable and transparent ML models, enabling users to understand how these models make predictions and identify potential sources of bias or error.

Overall, the use of ML in finance is a rapidly evolving area of research with the potential to transform the way financial markets are analyzed and understood. With careful attention to the challenges and opportunities presented by these techniques, ML has the potential to significantly benefit investors, traders, and policymakers while ensuring the responsible and ethical use of these powerful tools. So, ML is a powerful tool to predict the price of the financial markets.

Conflict of interest

The authors declare no conflict of interest.

References

1. Harries M, Horn K. Detecting concept drift in financial time series prediction using symbolic machine learning. In: Proceedings of The Eighth Australian Joint Conference on Artificial Intelligence; Canberra, Australia. pp. 91–98.
2. Cao L, Tay FEH. Financial forecasting using support vector machines. *Neural Computing & Applications* 2001; 10(2): 184–192.
3. Kim K. Financial time series forecasting using support vector machines. *Neurocomputing* 2003; 55(1–2): 307–319.
4. Huang W, Nakamori Y, Wang SY. Forecasting stock market movement direction with support vector machine. *Computers & Operations Research* 2005; 32(10): 2513–2522. doi: 10.1016/j.cor.2004.03.016
5. Kumar M, Thenmozhi M. Forecasting stock index movement: A comparison of support vector machines and random forest. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=876544# (accessed on 16 June 2023).
6. Ou P, Wang H. Prediction of stock market index movement by ten data mining techniques. *Modern Applied Science* 2009; 3(12): 28–42.
7. Majhi R, Panda G, Majhi B, Sahoo G. Efficient prediction of stock market indices using adaptive bacterial foraging optimization (ABFO) and BFO based techniques. *Expert systems with applications* 2009; 36(6): 10097–10104. doi: 10.1016/j.eswa.2009.01.012

8. Mohapatra P, Raj A, Patra TK. Indian stock market prediction using differential evolutionary neural network model. *International Journal of Electronics Communication and Computer Technology (IJECCT)* 2012; 2(4): 159–166.
9. Mao Y, Wei W, Wang B, Liu B. Correlating S&P 500 stocks with Twitter data. In: Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research; 12–16 August 2012; Beijing, China. pp. 69–72.
10. Siew HL, Nordin MJ. Regression techniques for the prediction of stock price trend. In: Proceedings of the 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE); 10–12 September 2012; Langkawi, Malaysia. pp. 1–5.
11. Shen S, Jiang H, Zhang T. Stock market forecasting using machine learning algorithms. Available online: https://cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearningAlgorithms.pdf (accessed on 16 June 2023).
12. Kazem A, Sharifi E, Hussain FK, et al. Support vector regression with chaos-based firefly algorithm for stock market price forecasting. *Applied Soft Computing* 2013; 13(2): 947–958. doi: 10.1016/j.asoc.2012.09.024
13. Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications* 2015; 42(1): 259–268. doi: 10.1016/j.eswa.2014.07.040
14. Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications* 2015; 42(4): 2162–2172. doi: 10.1016/j.eswa.2014.10.031
15. Rather AM, Agarwal A, Sastry VN. Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications* 2015; 42(6): 3234–3241. doi: 10.1016/j.eswa.2014.12.003
16. Xi Y, Peng H, Qin Y, et al. Bayesian analysis of heavy-tailed market microstructure model and its application in stock markets. *Mathematics and Computers in Simulation* 2015; 117: 141–153. doi: 10.1016/j.matcom.2015.06.006
17. Dash R, Dash PK. A hybrid stock trading framework integrating technical analysis with machine learning techniques. *The Journal of Finance and Data Science* 2016; 2(1): 42–57. doi: 10.1016/j.jfds.2016.03.002
18. Gerlein EA, McGinnity M, Belatreche A, Coleman S. Evaluating machine learning classification for financial trading: An empirical approach. *Expert Systems with Applications* 2016; 54: 193–207. doi: 10.1016/j.eswa.2016.01.018
19. Labiad B, Berrado A, Benabbou L. Machine learning techniques for short term stock movements classification for moroccan stock exchange. In: Proceedings of the 11th International Conference on Intelligent Systems: Theories and Applications (SITA 2016); 19–20 October 2016; Mohammedia, Morocco. pp. 1–6.
20. Usmani M, Adil SH, Raza K, Ali SSA. Stock market prediction using machine learning techniques. In: Proceedings of the 3rd International Conference on Computer and Information Sciences (ICCOINS 2016); 15–17 August 2016; Kuala Lumpur, Malaysia. pp. 322–327.
21. Tsantekidis A, Passalis N, Tefas A, et al. Using deep learning to detect price change indications in financial markets. In: Proceedings of the 2017 25th European signal processing conference (EUSIPCO); 28 August–2 September 2017; Kos, Greece. pp. 2511–2515.
22. Chen Y, Hao Y. A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications* 2017; 80; 340–355. doi: 10.1016/j.eswa.2017.02.044
23. Hitam NA, Ismail AR. Comparative performance of machine learning algorithms for cryptocurrency forecasting. *Indonesian Journal of Electrical Engineering and Computer Science* 2018; 11(3), 1121–1128. doi: 10.11591/ijeecs.v11.i3.pp1121-1128
24. Martin D, Póczos B, Hollifield B. Machine learning-aided modeling of fixed income instruments. In: Proceedings of the Thirty-second Conference on Neural Information Processing Systems; 2–8 December 2018; Montréal Canada.
25. Reddy VKS. Stock market prediction using machine learning. *International Research Journal of Engineering and Technology (IRJET)* 2018; 5(10): 1033–1035.
26. Ren R, Wu DD, Liu T. Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Systems Journal* 2019; 13(1): 760–770. doi: 10.1109/JSYST.2018.2794462
27. Lakshminarayanan SK, McCrae JP. A comparative study of SVM and LSTM deep learning algorithms for stock market prediction. In: Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science (AICS 2019); 12–13 July 2019; Wuhan, China. pp. 446–457.
28. Modi P, Shah S, Shah H. Big data analysis in stock market prediction. *International Journal of Engineering Research & Technology (IJERT)* 2019; 8(10). doi: 10.17577/IJERTV8IS100224

29. Mohan S, Mullapudi S, Sammeta S, et al. Stock price prediction using news sentiment analysis. In: Proceedings of the 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (Bigdata Service); 4–9 April 2019; Newark, CA, USA. pp. 205–208.
30. Reddy CV. Predicting the stock market index using stochastic time series ARIMA modelling: The sample of BSE and NSE. *Indian Journal of Finance* 2019; 13(8): 7–25. doi: 10.17010/ijf/2019/v13i8/146301
31. Zhong X, Enke D. Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation* 2019; 5(1): 1–20. doi: 10.1186/s40854-019-0138-0
32. Basak S, Kar S, Saha S, et al. Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance* 2019; 47: 552–567. doi: 10.1016/j.najef.2018.06.013
33. Michele A, Colin V, Santika DD. Mobilenet convolutional neural networks and support vector machines for palmprint recognition. *Procedia Computer Science* 2019; 157: 110–117. doi: 10.1016/j.procs.2019.08.147
34. Long W, Lu Z, Cui L. Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems* 2019; 164: 163–173. doi: 10.1016/j.knosys.2018.10.034
35. Berislav Ž, Hrvoje J. Forecasting stock market indices using machine learning algorithms. *Interdisciplinary Description of Complex Systems: INDECS* 2020; 18(4): 471–489. doi: 10.7906/indecs.18.4.6
36. Ghasemzadeha M, Mohammad-Karimi N, Ansari-Samani H. Machine learning algorithms for time series in financial markets. *Advances in Mathematical Finance and Applications* 2020; 5(4): 479–490. doi: 10.22034/AMFA.2020.674946
37. Khan W, Ghazanfar MA, Azam MA, et al. Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing* 2022; 13: 3433–3456. doi: 10.1007/s12652-020-01839-w
38. Kilimci ZH, Duvar R. An efficient word embedding and deep learning based model to forecast the direction of stock exchange market using twitter and financial news sites: A case of Istanbul stock exchange (bist 100). *IEEE Access* 2020; 8: 188186–188198. doi: 10.1109/ACCESS.2020.3029860
39. Obthong M, Tantisantiwong N, Jeamwatthanachai W, Wills G. A survey on machine learning for stock price prediction: Algorithms and techniques. In: Proceedings of 2nd International Conference on Finance, Economics, Management and IT Business; 5–6 May 2020; Online Streaming.
40. Chen W, Zhang H, Mehlawat MK, Jia L. Mean-variance portfolio optimization using machine learning—Based stock price prediction. *Applied Soft Computing* 2021; 100: 106943. doi: 10.1016/j.asoc.2020.106943
41. Janková Z. A bibliometric analysis of artificial intelligence technique in financial market. *Scientific Papers of the University of Pardubice. Series D, Faculty of Economics & Administration* 2021; 29(3). doi: 10.46585/sp29031268
42. Prasad A, Seetharaman A. Importance of machine learning in making investment decision in stock market. *Vikalpa: The Journal for Decision Makers* 2021; 46(4): 209–222. doi: 10.1177/02560909211059992
43. Subasi A, Amir F, Bagedo K, et al. Stock market prediction using machine learning. *Procedia Computer Science* 2021; 194: 173–179. doi: 10.1016/j.procs.2021.10.071
44. Tiwari S, Ramampiaro H, Langseth H. Machine learning in financial market surveillance: A survey. *IEEE Access* 2021; 9: 159734–159754. doi: 10.1109/ACCESS.2021.3130843
45. Beukel VB. *Influence of Machine Learning on Stock Pricing: A Meta-analysis* [Bachelor's thesis]. University of Twente; 2022.
46. Chen Q. Stock market prediction using machine learning. In: Qiu D, Jiao Y, Yeoh W (editors). *Atlantis Highlights in Intelligent Systems: Book 5*, Proceedings of the 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022); 25–27 March 2022; Wuhan, China. Atlantis Press; 2022. Volume 5, pp. 458–465.
47. Jishtu P, Prajapati H, Fiaidhi J. Prediction of the stock market based on machine learning and sentiment analysis. Available online: https://www.techrxiv.org/articles/preprint/Prediction_of_the_Stock_Market_Based_on_Machine_Learning_and_Sentiment_Analysis/21692852 (accessed on 16 June 2023).
48. Soni VK, Srivastava D. The use of supervised text classification techniques: A comprehensive study. In: Proceedings of the 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE 2022); 28–29 April 2022; Greater Noida, India.