Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

Machine Learning Approach for Financial Market Prediction: Review and Future Research

Mr. Rahul Maruti Dhokane[1]
(Research Scholar, Suresh Gyan Vihar University, Jaipur)

Prof. (Dr.) Om Prakash Sharma[2]
(Professor, GVSET, Suresh Gyan Vihar University, Jaipur.)

ABSTRACT- Financial market investment methods are difficult, requiring trust in the analysis of huge amounts of data. Nowadays, more research is being done to determine whether machine learning methods may enhance traditional methods of market prediction. Based on a survey of recent literature, the goal of this work is to identify the directions for future machine learning stock market prediction research. Peer-reviewed journal publications over the previous two decades are found using a step-by-step literature review technique, and research with similar methods and contexts is categorized. It has four categories: artificial neural network studies (ANN), support vector machine studies (SVM), Studies using hybrid or other artificial intelligence (AI) techniques , and genetic algorithms (GA) used in conjunction with other methods of analysis. Studies from every category are examined to detect patterns, unusual findings, restrictions, and areas that need more research. The overall results and methods for further research are presented in the final section.

Keywords- **artificial neural networks, support vector machines, genetic algorithms, research taxonomies, and stock market forecasting**

1. INTRODUCTION

The financial markets are used to create long-term wealth. In 2019-2020, the general value of the world's equities surpassed $85 trillion (Pound) [1]. Investors continuously try to find a decent strategy to spot sectors and companies for a decent return. Previously, investors used their experience to spot good quality companies, but today there's an enormous amount of information, so it's challenging to spot good sectors for investment. Statistical analysis of a company's data provides some guidelines. However, in recent years, investing firms have employed Artificial Intelligence (AI) algorithms to scan vast quantities of real-time economic and equity data for trends. These systems are used for investment decision-making. After a protracted period of using this technique, its performance is often reviewed and investors can identify which system is healthier for prediction than other techniques. The main focus of this study is to spot the longer-term direction for machine learning and deep learning financial market prediction supported by this literature review. The studies are accustomed to pointing out common results, uncommon results, limits, and areas that need more research. This could guide researchers and analysts in predicting stock exchange index values and stock movement. A scientific literature review technique has shed light on the use of machine learning (ML) in a wide range of computer applications and scientific fields. These are the papers that employed this approach and received a lot of citations. Wen et al. in 2012 published a scientific evaluation of the literature to evaluate the research of ML approaches to detect software development efforts (SDEE). The study chose articles between 1990 and 2009 [2]. The authors looked at each system's machine-learning approach, model estimation precision, model elements, and estimated context. The researcher's comments and guidelines for practitioners were based on insights into the current state of SDEE research that were supplied by this literature review.

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

Malhotra et al. in (2015) undertook a review of the literature for another IT-based project in which the prediction of software faults was undertaken [3]. For this, ML models were applied. To identify problematic modules and classes in the initial stages of the software development lifecycle, one technique is software fault prediction. The reviews of the articles included works released between 1991 and 2013. The article's appraisal was backed up by a comparison of their predicting abilities with those of various machine learning and statistical methods. In the end, seven categories were created from the publications. Overall, it was discovered that further work was needed to generate findings that might be used for ML-based software failure prediction. A set of recommendations for practitioners and researchers is produced in the study's conclusion.

Another study on artificial intelligence-based flood prediction systems was completed by Mosavi et al. in 2018, [4]. The modeling of intricate physical flood processes is a common application for machine learning approaches. With these models, hydrologists can forecast floods over the short and long term and find economically viable solutions that reduce risk, casualties, and property damage. Each article's ML methodology was examined to support its efficacy, speed, robustness, and correctness. The authors were prepared to offer recommendations for climate scientists and hydrologists when selecting the most straightforward machine learning approach for a prediction assignment, which they felt supported the review's findings.

One more study by Cabitza, Locoro, et al. applied ML techniques to examining orthopedic patients' data in 2018. The unit of medicine that focuses on preventing, identifying, and treating bone and muscle problems is called orthopedics. The authors found research on machine learning applications in orthopedics that had been written about for two decades; they collected data from orthopedic patients and applied ML methods to forecast the treatment based on symptoms [5].

This review article is categorized as follows: Section one provided a brief introduction. Section two provided the process used to find existing studies based on literature review and research questions. The research that supports the ML technique used to forecast stock trends and values for financial market indexes is then grouped into a framework (taxonomy) that is presented, supported by an examination of the studies that were identified. Section three introduces different methods to identify appropriate studies. Each study in each category is summarized and reviewed to find any problems. The final section provides a summary of the overall study goal, which focuses on providing directions for further study and ideas for research progress.

2. LITERATURE REVIEW

A. RESEARCH QUESTION

The systematic literature review is the response to one or more study questions. The research questions for this literature review are developed and provided in this section. The research questions for this systematic literature review are listed below.

RQ 1: "How can stock market forecasting assist investors in making the best choice? Why is it necessary?"

More and more people are becoming interested in the securities market. To generate a profit, they take risks with their money. The risk of losing the money invested or experiencing a loss is the possibility when investors invest money in the financial market with the wrong decision. If a person invests in a stock whose future value is expected to decline, they will suffer a loss. The estimation of the stock's longer-term worth must meet certain criteria for that reason financial market forecasting is a very important task.

RQ 2: "What are the various characteristics and variables that can be used to forecast the stock market?"

Searching for an ML model which can accurately predict stock prices is one of the most hotly debated subjects. A prior data set is required to train the

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

machine learning algorithm. This data set may include information in the form of different variables or attributes. A machine learning algorithm needs to be trained using a knowledge set to be able to anticipate the long-term value of a stock. Several attributes may be used to support this data collection. Researchers have used several variables to train their models. However, a few characteristics are frequently used. The stock price could also be impacted by some news, events, policy changes, etc.

RQ 3: "What varieties of machine learning forecasting algorithms do researchers employ to forecast the stock market?"

Linear Regression
Artificial Neural Network
Random Forest
K-Nearest Neighbor
Support Vector Machine (SVM)
Genetic Algorithm

B. Studies which are predicting financial market values with Artificial Neural Networks.

Studies that specialize in stock prediction using artificial neural networks are included in the first set of articles. Computer models known as ANNs are underpinned by biological neural networks. There are three layers of nodes within the network, the input layer comes first, the middle hidden layer is present and the output layer comes last. As the linked nodes learn supported related examples and attempt to lessen the magnitude of prediction error, signals are sent (propagated) throughout the network. Weights are modified for the signals between connected nodes because the system is operating to improve its performance. The ensuing information may briefly summarize the distinctive research focus and conclusions of each ANN-related work. Using information from various international stock markets, Jasic et al. in 2004 presented a systematic neural network approach to estimate daily exchange index results. The primary goal is to pursue lucrative trading and generate profit [6].

Researchers provide a method for short-term security market index return prediction based on univariate neural networks with untransformed data inputs. The German DAX Index, the Japanese TOPIX index, the Standard & Poor's 500 Index (S&P 500), and the London Financial Times Exchange Index are all used in the study. The S&P 500, DAX, and FTSE Indexes' sample data spans between January 1965, and November 1999. Since data for earlier years weren't accessible, the sample data for TOPIX only includes the period from January 1969, to November 1999. When applied to the S&P 500 and DAX indices, the neural network's prediction performance is measured against a benchmark linear autoregressive model and the prediction improvement is validated.

David Enke et al. in the year (2005) assess the prediction correlations for various financial and economic factors using the machine learning information gain approach [7]. A ranking of the model variables is created by computing the data gained for each one. A cutoff point is decided upon to choose only the most significant relevant variables to be kept in the forecasting models. It is investigated if neural network level estimation and classification models can accurately predict future values. One more method i.e. cross-check method is also implemented to improve the generalizability of different models. Using S&P data for 24 years, between March 1976 to December 1999, the models are contrasted. The findings demonstrate that, compared to holding strategies, opposing neural network models, and simple regression models, trading strategies led by classification models produce superior risk-adjusted profits.

Liao et al. in 2010 introduced the following study using a stochastic time-efficient neural network model to identify the links between various financial and economic factors that can be used to forecast outcomes. The weights are assigned to the stocks as per the relativity of the presumption that investors select their investment positions by reviewing past stock exchange data [8]. The influence of the information on the predictive model is stronger the closer the historical data is to this. By using a

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

numerical experiment supported by data from each trading day over 18 years, between December 1990, to June 2008, the effectiveness of the model is examined. A few of the financial markets from which the data is obtained include the NASDAQ index, Dow Jones Industrial Average index, and S&P 500 index. The model's forecasting ability is assessed using a variety of volatility traits.

To expand understanding of ANN market prediction, Kim, K. J., et al. 2004 investigated different model input parameters found in nine published research [9]. They seek the most significant input factors that improve the precision of model prediction. They found that whereas microeconomic factors are often used to forecast the values of securities market indexes, the bunch of ML techniques employ technical characteristics rather than fundamental variables to predict a specific stock price. Additionally, when compared to using only one type of input variable, hybridized parameters yield greater results.

Deep learning networks for financial market analysis and forecasting are examined by Chong et al. In 2017. Deep learning is not relying on historical data, it will calculate the characteristics based on calculations and after that, it will make decisions [10].

C. **Studies which prefer Support Vector Machines to Analyze Financial Markets**

Studies forming securities market forecasts using support vector machines (SVMs) are included in the second category of papers. Through example classification, SVMs provide an alternative approach to ANNs for increasing stock market prediction accuracy. Supervised learning is used in the technique. Examples used in training are labeled as belonging to one category or another. To create the widest possible gap between the categories, an SVM model portrays the examples as points in an exceeding space. For instance, according to Schumaker et al. in the year 2010, in the context of predicting exchange prices, SVM may be a machine learning method that may classify the direction of future stock prices (uptrend or downtrend) [11]. To predict the uptrend or downtrend of stock markets, A support vector machine and a hybrid feature selection method were linked in Ming-Chi Lee's 2009 prediction model [12]. The advantages of filter methods and wrapper methods are combined in this suggested hybrid, the F-score and supported sequential forward search feature selection techniques (F-SSFS), to select the most useful feature subset from the starting feature set. Performance is compared with a backpropagation neural network (BPNN) and information gain, correlation-based feature selection, and symmetrical uncertainty using paired t-tests are three often used feature selection techniques in order to evaluate the performance of this F-SSFS and SVM-based model for prediction. The major aim of the study is to forecast the direction of the NASDAQ index using data from commodities, currencies, and other financial market indices between November 2001, and November 2007. It has been found that the SVM performs better in forecasting stock trends than the BPN. An SVM was employed in addition to textual analysis in Schumaker R. P. et al. in 2009, who conducted an innovative study to look at how news stories affect stock prices [13]. They developed a technique in which financial news is analyzed for prediction by applying a machine learning approach. The data of NEWS between October 2005, and November 2005, examined many financial news articles and stock quotes about stocks in the S&P 500. Twenty minutes after an article was published, they determined a discrete stock value. They demonstrated using a simulated trading engine that the model which took into account both the terms of the article and the stock value at the time of the article's release provided the improved return. They did this by using an SVM derivative designed especially for models with many stock-specific variables and discrete numeric forecasting.

Yeh, C. Y., Et al., in 2011 outlined issues that can occur when utilizing kernel function hyperparameters to predict exchange prices when employing support vector regression. A parameter

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

whose value is ready before the instructional process begins is typically considered a hyperparameter. Their solution will combine the advantages of several hyperparameter settings, potentially improving system performance [14]. They develop a two-stage multiple-kernel learning method by fusing gradient projection with sequential minimal optimization. According to experimental results using datasets from the Taiwan Capitalization Weighted index, the new strategy beats previous methods.

With the use of daily stock price records from October 2002 to March 2005, the model was evaluated, trained, and validated. The authors of the study that follows find that marketplaces can have different features depending on where they are. They evaluate a machine learning model to do this utilizing information from the National Stock Exchange (NSE) of India for a time period of 2007 to 2010.

D. Studies analyzing financial markets using Genetic Algorithms (GA) and other supporting methods

Systems relying entirely on ANNs or SVMs, as described in the first two study methodologies, can improve financial market value prediction to a certain extent, but there seems to be a growing interest over time in trying to further enhance outcomes using multiple techniques and approaches. One alternative machine learning technique typically mixes genetic algorithms with either ANNs or SVMs to overcome the limitations of a single approach. One type of evolutionary algorithm is a genetic algorithm. Holland et al. in 1992 introduced, the selection of randomly generated problem solutions serves as the starting point for the evolutionary process. In every iterative generation, an objective function evaluates the fitness of every solution. A new generation of solutions is created by combining the higher fitness solutions with additional high fitness solutions (survival of the fittest) [15].

When parent solutions combine with a child a substitute child solution is created that incorporates certain traits from both parent solutions. For achieving better results this process will repeat. Systems combining ANNs with SVMs, and GAs are developed in the research. In their initial investigation on this topic in 2000, Kim, K. J. et al. [16] Introduced artificial neural networks (ANNs) that use genetic algorithms to choose connection weights and discretize features to predict stock value indicators. For network training, feature subset selection, and topology optimization, previous research papers combined GAs and ANNs. But in the majority of those experiments, the GA is hardly ever used to enhance the training algorithm itself. In this study, GA is employed to mitigate the problem in the feature space while also enhancing the training algorithm. Technical indicators and, consequently, the direction of change within the daily Korea Stock Index Number (KOSPI) between January 1989 to December 1998 are the research data used in this study. According to experimental findings, the feature discretization model's GA technique is more effective than the other two traditional approaches.

The same two machine learning techniques used in the earlier study were also supported by Kim, K. J., & Lee, W. B. in 2004 [9]. They contrast two traditional artificial neural network (ANN) approaches with a feature transformation method using a GA. The educational and generalizability of ANNs for market prediction are improved with the use of the GA. The study data covers 2,347 trading days between January 1991 to December 1998 and contains technical indicators. The experimental findings, according to the authors, show that the suggested method reduces the feature space's dimensionality and removes unnecessary variables for predicting the financial market.

A novel hybrid system was created by Kim, M. J., Min, S. H., & Han, I. in 2006 to forecast the values of security market indexes using ANN and GA [17]. The idea behind their method is various classifier combinations, where different classifiers attempt to provide a solution to the same issue before combining their findings to lower estimation errors and boost classification accuracy overall. They

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

discovered that there are restrictions when using this approach to resolve business issues because it is challenging to fully explain the outcomes produced by ML-driven classifiers due to the complexity of the issue. This paper suggests a method for infusing human subjectivity in problem-solving into the output of quantitative models. Because of the computer-driven classifier, the authors employ a three-layered backpropagation neural network. Gene-based algorithms Classifiers from three different sources: experts, machine learning, and users are combined using GA. The Korea Stock Indicator (KOSPI) provided training, testing, and validation data for 573 weeks between January 1990 and December 2001.

The efficiency of a hybrid genetic algorithm and artificial neural network technology for stock exchange prediction was introduced by Kim, H. J., & Shin, K. S. in 2007 [18]. The study makes use of two distinct networks: time-delay neural networks and adaptive time-delay neural networks (ATNNs) (TDNNs). In addition to many heuristics and statistical techniques, a broad strategy based on trial and error is suggested to estimate the components of the ATNN and TDNN designs. A Genetic Algorithm is used in the models for optimization purposes. The daily Korea Stock Price Level 200 (KOSPI 200) between January 1997 through December 1999 served as the source of research data for this study. The outcomes demonstrate that the integrated strategy is more effective than the quality ATNN, TDNN, and recurrent neural network (RNN).

To identify trade patterns in this hybrid GA, a developing approach using the least squares support vector machine (LS-SVM) learning paradigm with a mixed kernel is proposed and studied by Yu, L., Chen, H., et al. in 2008 [19] pertaining to SVM. In the suggested learning paradigm, the input features for LS-SVM learning are initially chosen using a genetic algorithm. The LSSVM's parameter settings are then optimized using a different GA. Finally, the direction of movement of the financial market is predicted using historical data series using the growing LS SVM learning paradigm using a mixed kernel, the best feature subset, and the best parameters. Data from the S&P 500, the stock market's Industrial Average, and the New York Stock Exchange Index are used in testing for evaluation purposes. With a total of 958 observations, the whole monthly value data set spans the period between January 1926 to December 2005. The suggested evolving LS-SVM can provide some forecasting models that are relatively easier to understand since they require fewer predictive features and are more effective than previous parameter optimization techniques, according to experimental data.

Chiu et al. in (2009) suggest using an SVM and a dynamic fuzzy model to identify stock exchange volatility. The fuzzy model uses factors with varying levels of influence to integrate input variables [20]. A GA dynamically modifies each input variable's degree of influence. The SVM is then used to predict the dynamics of the securities market. A multi-period experiment is used to simulate exchange volatility. The study's 58 input parameters comprise macroeconomic data, exchange technical indicators, and technical indicators for the commodities market. They contrast the new integrated model's performance with that of conventional forecasting techniques. Macroeconomic indicators for the period between January 2003 to December 2004 come from the Ministry of Economic Affairs, ROC, while stock and commodity exchange figures are from the Taiwan Stock Market Corporation. The results of the experiments show that the model is superior to other prediction methods in terms of accuracy.

E. Studies Analyzing Stock Markets with Hybrid or Other AI Techniques

The most popular methods for addressing the issue of security market prediction include ANNs, SVMs, and multi-method GA approaches. This final category includes works that have applied additional distinctive or multi-method AI techniques to this area

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

of problems. Rule-based expert systems have been used for many years to give pertinent decision-makers domain-specific information.

A candlestick chart analysis expert system was created by Lee and Jo in 1999 [21] to forecast the best time to enter the securities market. The expert system has patterns and rules that could foresee future changes in stock prices. Five different forms of price movements are used to characterize defined patterns: patterns of growing, decreasing, neutrality, trend continuance, and trend reversal. Results suggest that the model they developed may provide indicators to assist investors in boosting the returns on their stock investments. The constructed knowledge domain was demonstrated through experiments utilizing data from a sample of stocks posted on the Korean exchange between January 1992 and June 1997 to be independent of both time and field. Another crucial stock-related financial choice is asset allocation. However, this topic has gotten less attention in machine learning studies.

In a framework that is heavily reliant on reinforcement learning, O, Lee, et al. in (2006) [22] presented a replacement stock trading technique that is linked to dynamic asset allocation. The proposed asset allocation technique, known as "meta policy" (MP), is designed to make use of the temporal data from both stock recommendations and the stock fund's ratio to assets. The Meta Policy (MP) is designed using a learning agent environment within the reinforcement learning framework. The purpose of the following study is to combine various machine-learning approaches for the forecast of the financial market. Stock prices should exhibit a stochastic process pattern according to the efficient market theory, indicating that the market cannot be predicted with greater than a 50% degree of accuracy.

To demonstrate that not all eras are equally random, Qian et al. in 2007 [23] investigated the predictability of the stock index, the Industrial Average index. They chose a period of high predictability using the Hurst exponent and discovered that the most beneficial study period was between June 5, 1969, to June 5, 1973. (1000 trading days). The following three inductive machine learning classifiers were used to forecast financial market prices: k-nearest neighbor, a decision tree, and a synthetic neural network. Because the models worked together well, the resultant prediction accuracy was 65%, which was higher than random. The study that comes after uses the widest variety of machine learning methods. Data mining techniques have produced reliable predictions of future stock price movements, but traders have come to realize that they need to combine different prediction techniques to get a deeper understanding of the securities market over the long run.

In this 2009 study by Ou et al. different data processing approaches are applied to forecast price movement inside the Hang Seng index on the Hong Kong exchange [24]. Tree-based classification, naive Bayes based on kernel estimate, logit model, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and method of least squares support vector machine (LS-SVM) are among the techniques. Using data between January 3, 2000, to December 29, 2006, By placing bets on five predictors, they looked at the daily variation in closing prices within the Hang Seng index. According to experimental findings, the SVM and LS-SVM outperform the opposing models. In terms of various hit rate and error rate criteria, SVM outperforms LS-SVM, whereas, for out-of-sample prediction, LS-SVM outperforms SVM. In order to outperform traditional linear and nonlinear approaches for exchange prediction, a number of neural network models and hybrid models are provided. The majority of ANN models in this field have certain limitations.

A hybrid neural network that uses generalized autoregressive conditional heteroscedasticity (GARCH) to extract additional input variables, a dynamic artificial neural network (DANN), a multi-layer perceptron (MLP), and others are all evaluated by Guresen, E., et al., in 2011 [25]. Daily NASDAQ index readings between October 7, 2008, to June 26, 2009, are used to compare various

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

approaches. One discovery is that the straightforward MLP appears to be the most effective and practical ANN architecture. This study's primary objective is to forecast Asian stock markets, which are very unpredictable and active. An Asian stock market statistical prediction model presented in a work by Dai, et al. in 2012 has been developed which contains nonlinear independent component analysis (NLICA) and artificial neural networks (ANN) [26]. In the absence of pertinent data mixing mechanisms, NLICA might be a cutting-edge feature extraction technique for distinguishing distinct sources from nonlinear mixture data that has been observed. In the suggested method, the primary goal of NLICA is to convert the input space, which consists of the original statistical data, into a feature space, which consists of discrete components that represent the data's underlying information.

To build the neural network's prediction model, the independent components are then employed as input variables. The performance of the proposed method is illustrated using data between 2 February 2004 through March 3, 2009, from the Nikkei 225 closing index and Shanghai type B indexed share closing price. According to experimental findings, the suggested forecasting model not only beats the three comparison approaches but also increases prediction accuracy when using a neural network approach. In one study, Patel, Shah, et al. analyze the Indian security market with four prediction models (ANN, SVM, random forest, and naive Bayes) as well as two model input approaches in 2015 [27]. While the second strategy focuses on capturing these technical characteristics as trend deterministic data, the principal approach for computer files entails computing different technical criteria utilizing stock market data (open, high, low, and shutting prices). They examine the accuracy of each forecasting model for each of the two input techniques between January 2003 and December 2012, data from 50 stocks listed in different indices like BSE, NSE, and S&P 500 were used to conduct the analysis. The computer file technique is less efficient than random forest. The performance of all prediction models increases using a rule set and a computationally efficient functional link artificial neural network (CEFLANN), Dash et al. 2016 developed a wholly original decision web [28]. They see the choice of whether to buy, hold, or sell stocks as a classification problem with three possible outcomes. The BRT algorithm seeks to optimize the forecast error's predictive power and endogenously chooses the predictor variables used to mimic the forecasters' knowledge base. The prediction error's potential nonlinear reliance on the predictor variables as well as their interdependencies are both taken into account by the BRT algorithm. The S&P 500 index estimates from three groups for the years 1992 to 2014 are included in the study's data. According to their main discovery, given the collection of predictor variables used in this analysis, longer-term forecasts are contrary to the rational expectations hypothesis (REH), according to the evidence. The results from three different forecaster groups support the majority of the conclusions. A method to forecast the daily return direction of a group of stocks was presented by Zhong et al. in 2019 [29]. To predict the daily direction of upcoming stock exchange index returns, deep neural networks (DNNs) and conventional artificial neural networks (ANNs) are used to the entire preprocessed but untransformed dataset together with two other datasets (PCA). Simulation results demonstrate that compared to other hybrid machine learning algorithms or the entire untranslated dataset, DNNs using two PCA-represented datasets considerably improve classification accuracy. The trading strategies assisted by the DNN classification method outperformed the others examined, including a comparison against two common benchmarks on PCA-represented data. The dataset used in this study includes 58 financial and economic indicators as input features, as well as the SPDR S&P 500 ETFs' daily direction (uptrend or downtrend) as output characteristics. From June 1, 2003, to May 31, 2013,

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

there were 2518 trade days, which constitute the daily statistics.

3. METHOD FOR DETERMINING APPROPRIATE STUDIES

Each researcher who took part in this study independently searched for peer-reviewed journal articles that used a particular type of machine learning to forecast a stock market outcome. Using Google Scholar, EBSCO, and EconLit, articles were found. Only studies over the last two decades (1999 to 2020) were used to compile the final list in order to identify conclusions that are relevant to the current technical environment. ML approaches are used to decide whether index value would increase or decrease in the long term. An initial set of 40 pertinent articles was found after the elimination of repeated documents.

Many pieces of research were taken off the list because they only attempted to forecast the value of a single stock. As an illustration, A genetic algorithm (GA) was developed in one study by Kumar et al., in 2011 to forecast the value of two particular stocks: Infosys and Tata Consultancy Services [30]. To provide a proportional sample of research studies in this field, 25 studies were eventually added to the final list. The list does contain enough coverage to allow for conclusions and suggestions, but it is not meant to be an exhaustive list of all articles that are linked. Each researcher then went through each manuscript to look for clusters of linked studies that each utilized a single machine learning technique or studies that used multiple methods. The resulting machine learning security market research taxonomy is the end outcome. Each of the following four categories best describes one of the articles:

(1) Research into artificial neural networks
(2) Study of support vector machines
(3) Research combining genetic algorithms with other techniques.
(4) Research using hybrid or alternative computer science methods

TAXONOMY OF MACHINE LEARNING STOCK MARKET PREDICTION STUDIES

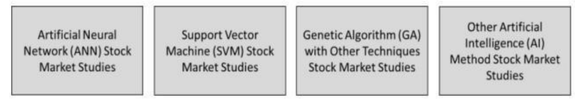

Figure 3.1. Research Taxonomy Study Machine Learning Stock Market Prediction.

The specific publications that fall under each study taxonomy category and are specialized in their distinct models, datasets, and contributions are summarized in the section that follows. The table has a complete list of the studies that have been reviewed. Before describing the relevant findings, a summary of each machine-learning approach is also provided.

TABLE 3.1

List of Machine Learning Methods

Authors Details	Machine Learning Techniques Used.	Research Publication Year	Datasets and Time Period in Years
Enke, Thawornwong	ANN	2005	S&P 500, 1976 to1999
Liao, Wang	ANN	2010	Shanghai Stock Exchange 1990 to 2008
Jasic, Wood	ANN	2004	DAX, TOPIX, S&P 500, and FTSE 1965 to 1999
Kim, K. J., & Lee, W. B.	ANN	2004	Study reviews of nine ANN studies
Kim, Han	GA with ANN	2000	Korea KOSPI index 1989 to 1998
Lee	SVM	2009	NASDAQ index 2001 to 2007

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

Yeh, Huang and Lee	SVM	2011	Taiwan Stock index from 2002 to 2005
Das, Padhy	SVM	2012	NSE index of India Limited 2007 to 2010
Schumaker, Chen	SVM	2009	Companies listed in the S&P 500 in 2005
Kim and Lee	GA with ANN	2004	KOSPI index 1991 to 1998
Kim and Shin	GA with ANN	2007	KOSPI 200 index 1997 to 1999
Yu, Chen, Wang, and Lai	GA in combination with SVM	2008	S&P 500, NYSE Index from 1926 to 2005
Chiu, Chen	GA with SVM	2009	Taiwan Stock Exchange 2003 to 2004
Qian and Rasheed	ANN with decision tree and k-nearest neighbor	2007	DJIA 1969 to 1973
Chong, Han and Park	ANN	2017	KOSPI stock index 2010 to 2014
Lee and Jo	Candlesticks.	1999	Stocks from the Korean stock market 1992 to 1997
Dai, Wu, and Lu	NLICA in combination with ANN	2012	Nikkei 225, Shanghai Type B-share closing index 2004 to 2009
Patel, Thakkar, Kotecha, and Shah	SVM, ANN, and naïve Bayes with supporting data.	2015	CNX Nifty and S&P, BSE Sensex index 2003 to 2012
Dash and Dash	ANN and ruleset.	2016	SENSEX index and S&P 500 index, 2010 to 2014
Jangmin, Lee, and Zhang	MP	2006	KOSPI index from 1998 to 2003
Pierdzioch, Risse	Regression	2018	S&P 500 index from 1992 to 2014
Zhong, Enke	Deep neural networks (DNNs) and traditional ANN	2019	S&P 500, 2003 to 2013
Kim, Min and Han	GA with ANN	2006	KOSPI index 1990 to 2001

4. COMPARISON AND FUTURE RESEARCH DIRECTIONS

This study's primary goal is to outline future machine learning financial market prediction research directions based on an analysis of the literature currently available. This topic has drawn different conclusions regarding the state of our current understanding and often the taxonomic categories already introduced in Machine Learning based systems, and discoveries described in each selected article. Machine Learning techniques and the relevant prediction problems have a good link. It is associated with a work-technology fit, where system performance is controlled by effective matching between functions and technologies (Goodhue et al. in 1995) [31]. The best use of artificial neural networks (ANN) is for forecasting numerical exchange index values. Support vector machines are a good choice when determining

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

whether the stock exchange index is expected to be in an uptrend or downtrend. To generate the simplest returns, genetic algorithms (GA) estimate which stocks to include in a particular portfolio or find higher-quality system inputs. While every study showed that the methods could be used successfully, there are restrictions on how the methods can be used.

The use of hybrid machine learning techniques could be one way to get around some of these restrictions. The problem is that, at a certain point, the systems get too complicated and are no longer practical. Future investigations may introduce this as a theoretical and practical issue. The generalizability of results needs to be enhanced, which is the next finding from this analysis of prior research. Most studies just use one market and one era to evaluate their ML systems, not taking into account whether the system will work well in different environments. Three improvements will be made to the experimental system evaluation. The first findings from Asian stock markets are used to support several studies. Even testing of these systems for the US or European markets might be done in the same time frame. Second, the systems' performance may be examined across a range of market situations using data from periods when markets are rising or falling. For instance, could a method successfully forecast market values in the US throughout the 2008–2009 financial crisis as well as the most recent market expansion from 2018–2019? Are systems capable of forecasting market expansion as well as market contraction? Finally, the suggested methodologies may be used to compare the predicted performance of stock exchange indices that comprise both small and large companies. Are systems reliable in a range of risk and volatility situations? Any of those improvements to the scientific process will make a greater contribution to research and practice. After some thought, the final set of findings was similarly clear. The inputs, algorithms, and performance metrics of ML systems must have a stronger scope at their core. If this is not the case, then the outcomes might be arbitrary and useless. Too many studies disregard the enormous body of monetary theory that has been established over the last few centuries in favor of procedures.

REFERENCES

[1] Pound, J. (2019, December 24). Global stock markets gained $17 trillion in value in 2019. Retrieved from https://www.cnbc.com/2019/12/24/global-stock-markets-gained-17-trillion-in-value-in-2019.html. Year 2019.

[2] Wen, J., Li, S., Lin, Z., Hu, Y., & Huang, C. Systematic literature review of machine learning based software development effort estimation models. Information and Software Technology, 54(1), 41-59 Year 2012

[3] Ruchika Malhotra, A systematic review of machine learning techniques for software fault prediction, Applied Soft Computing, Volume 27, Pages 504-518, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2014.11.023., Year 2015

[4] Mosavi, A.; Ozturk, P.; Chau, K.-w. Flood Prediction Using Machine Learning Models: Literature Review. Water 2018, 10, 1536. https://doi.org/10.3390/w10111536 Year 2018

[5] Cabitza, F., Locoro, A., & Banfi, G. Machine learning in orthopedics: a literature review. Frontiers in Bioengineering and Biotechnology, 6, 75 .https://doi.org/10.3389/fbioe.2018.00075 Year 2018.

[6] Teo Jasic & Douglas Wood The profitability of daily stock market indices trades based on neural network predictions: case study for the S&P 500, the DAX, the TOPIX and the FTSE in the period 1965–1999, Applied Financial Economics, 14:4, 285-297, DOI: 10.1080/0960310042000201228, Year 2004.

[7] David Enke, Suraphan Thawornwong, The use of data mining and neural networks for forecasting stock market returns,Expert Systems with Applications, Volume 29, Issue 4, Pages 927-940, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2005.06.024. Year 2005.

[8] Zhe Liao, Jun Wang, Forecasting model of global stock index by stochastic time effective neural network, Expert Systems with Applications, Volume 37, Issue 1, Pages 834-841, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2009.05.086. Year 2010.

[9] Kim, Kj., Lee, W.B. Stock market prediction using artificial neural networks with optimal feature

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

transformation. Neural Comput & Applic 13, 255–260. https://doi.org/10.1007/s00521-004-0428-x Year 2004.

[10] Chong, Eunsuk et al. "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies." Expert Syst. Appl. 83: 187-205. Year 2017.

[11] R. P. Schumaker and H. Chen, "A Discrete Stock Price Prediction Engine Based on Financial News," in Computer, vol. 43, no. 1, pp. 51-56, doi: 10.1109/MC.2010.2. Year 2010

[12] Ming-Chi Lee, Using support vector machine with a hybrid feature selection method to the stock trend prediction, Expert Systems with Applications, Volume 36, Issue 8, Pages 10896-10904, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2009.02.038. Year 2009.

[13] Schumaker, R. P., & Chen, H. Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Transactions on Information Systems (TOIS), 27(2), 12. Year 2009.

[14] Yeh, C. Y., Huang, C. W., & Lee, S. J. A multiple-kernel support vector regression approach for stock market price forecasting. Expert Systems with Applications, 38(3), 2177-2186. Year 2011.

[15] Holland, J. H. (1992). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press. Year 1992.

[16] Kyoung-jae Kim, Ingoo Han, Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index, Expert Systems with Applications, Volume 19, Issue 2, Pages 125-132, ISSN 0957-4174, https://doi.org/10.1016/S0957-4174(00)00027-0 Year 2000.

[17] Myoung-Jong Kim, Sung-Hwan Min, Ingoo Han, An evolutionary approach to the combination of multiple classifiers to predict a stock price index, Expert Systems with Applications, Volume 31, Issue 2, Pages 241-247, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2005.09.020. Year 2006.

[18] Hyun-jung Kim, Kyung-shik Shin, A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets, Applied Soft Computing, Volume 7, Issue 2,Pages 569-576, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2006.03.004 Year 2007.

[19] L. Yu, H. Chen, S. Wang and K. K. Lai, "Evolving Least Squares Support Vector Machines for Stock Market Trend Mining," in IEEE Transactions on Evolutionary Computation, vol. 13, no. 1, pp. 87-102, DOI: 10.1109/TEVC.2008.928176. Year 2009.

[20] Deng-Yiv Chiu, Ping-Jie Chen, Dynamically exploring internal mechanism of stock market by fuzzy-based support vector machines with high dimension input space and genetic algorithm,Expert Systems with Applications, Volume 36, Issue 2, Part 1, Pages 1240-1248, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2007.11.022. Year 2009.

[21] K.H. Lee, G.S. Jo, Expert system for predicting stock market timing using a candlestick chart, Expert Systems with Applications, Volume 16, Issue 4, Pages 357-364, ISSN 0957-4174, https://doi.org/10.1016/S0957-4174(99)00011-1. Year 1999.

[22] Jangmin, O, Jongwoo Lee, Jae Won Lee and Byoung-Tak Zhang. "Adaptive stock trading with dynamic asset allocation using reinforcement learning." Inf. Sci. 176: 2121-2147. Year 2006.

[23] Qian, B., Rasheed, K. Stock market prediction with multiple classifiers. Appl Intell 26, 25–33. https://doi.org/10.1007/s10489-006-0001-7 Year 2007.

[24] Ou, P., & Wang, H. Prediction of stock market index movement by ten data mining techniques. Modern Applied Science, 3(12), 28-42. Year 2009.

[25] Erkam Guresen, Gulgun Kayakutlu, Tugrul U. Daim, Using artificial neural network models in stock market index prediction, Expert Systems with Applications, Volume 38, Issue 8, Pages 10389-10397, ISSN0957-4174, https://doi.org/10.1016/j.eswa.2011.02.068. Year 2011.

[26] Dai, W., Wu, J. Y., & Lu, C. J. Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. Expert systems with applications, 39(4), 4444-4452. Year 2012

[27] Jigar Patel, Sahil Shah, Priyank Thakkar, K Kotecha, Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques, Expert Systems with Applications, Volume 42, Issue 1, Pages 259-268, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2014.07.040. Year 2015.

[28] Rajashree Dash, Pradipta Kishore Dash, A hybrid stock trading framework integrating technical analysis

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi-Annual Journal), Volume 8, Issue 2, 2022, pp 19-31
ISSN: 2395-0196

with machine learning techniques, The Journal of Finance and Data Science, Volume 2, Issue 1, Pages 42-57, ISSN2405-9188,https://doi.org/10.1016/j.jfds.2016.03.002 Year 2016.

[29] Zhong, X., Enke, D. Predicting the daily return direction of the stock market using hybrid machine learning algorithms. Financ Innov 5, 24. https://doi.org/10.1186/s40854-019-0138-0 Year 2019.

[30] Kumar, L. Kuladeep et al. “A Hybrid Machine Learning System for Stock Market Forecasting.” Journal of International Technology and Information Management 20. Year 2011.

[31] Goodhue, Dale L., and Ronald L. Thompson. "Task-technology fit and individual performance." MIS quarterly (1995): 213-236. Year 1995.