

The Stock Exchange Prediction using Machine Learning Techniques: A Comprehensive and Systematic Literature Review

¹Rico Bayu Wiranata, ²Arif Djunaidy

^{1,2}Department of Information System, Intelligent Electrical and Informatics Technology, Sepuluh
Nopember Institute of Technology, Surabaya, 60111, Indonesia

E-mail: ¹rico.19052@mhs.its.ac.id ²adjunaidy@is.its.ac.id

Abstract

This literature review identifies and analyzes research topic trends, types of data sets, learning algorithm, methods improvements, and frameworks used in stock exchange prediction. A total of 81 studies were investigated, which were published regarding stock predictions in the period January 2015 to June 2020 which took into account the inclusion and exclusion criteria. The literature review methodology is carried out in three major phases: review planning, implementation, and report preparation, in nine steps from defining systematic review requirements to presentation of results. Estimation or regression, clustering, association, classification, and preprocessing analysis of data sets are the five main focuses revealed in the main study of stock prediction research. The classification method gets a share of 35.80% from related studies, the estimation method is 56.79%, data analytics is 4.94%, the rest is clustering and association is 1.23%. Furthermore, the use of the technical indicator data set is 74.07%, the rest are combinations of datasets. To develop a stock prediction model 48 different methods have been applied, 9 of the most widely applied methods were identified. The best method in terms of accuracy and also small error rate such as SVM, DNN, CNN, RNN, LSTM, bagging ensembles such as RF, boosting ensembles such as XGBoost, ensemble majority vote and the meta-learner approach is ensemble Stacking. Several techniques are proposed to improve prediction accuracy by combining several methods, using boosting algorithms, adding feature selection and using parameter and hyper-parameter optimization.

Keywords: *systematic literature review, stock market prediction method, stock prediction dataset, prediction method improvement, prediction framework, machine learning*

1. Introduction

Research on financial time series in the last five years has achieved rapid development, based on intelligent learning algorithms, such as forecasting regression [1] and classification [2]. Including, this result has attracted widespread attention for economists, investors, investment managers, and even data scientists. For example, [3], [4], [5] proposed an algorithm to predict the stock exchange using the decision tree method in a random forest or ensemble bagging decision tree. [6], [7], [8], [9] studied stock trading through deep learning such as deep neural network, long short-term memory, recurrent neural network, and convolutional neural network.

For superior results, several studies have used one or a combination of various types of comprehensive data sets including economics, politics, stock trading time series, company fundamentals, news sentiment, social media, and

commodity trading movements. For example, [10] used indicator technical data to predict stocks on the China Shanghai Stock Exchange market, then [11] proposed stock predictions using a combination of types of technical and macroeconomic indicator data sets. Combining technical indicators and news sentiment through text mining techniques such as [12], [13], [14], [4], [15], research [16], [3], [17] also found that the company's fundamentals had a positive impact on changes in the company's stock price.

Many types of data sets, methodologies, method modifications, combinations of methods and stock market prediction frameworks published are different and complex, making a helicopter view of the status of existing stock research blurry and disappearing. In addition, for data scientists who want to develop stock prediction models, it is quite difficult to comprehensively look at the gaps from recent research. For example, in a systematic literature

review [18] only explaining modeling techniques and types of input data sets only considers technical analysis and fundamental analysis. [19] only describe methods, types of data sets, and performance evaluation metrics and only on classification topics in their systematic literature review. [20] in their systematic literature review only describes the types of data sets, input variables, methods, performance evaluation metrics. However, the above systematic literature review has not sufficed the needs of data scientists or future researchers for gap analysis and developing stock predictions.

Therefore, it is important to present a complete systematic literature review that has not been done before, such as all kinds of potential topics in stock prediction such as clustering, association, classification, regression, forecasting, and dataset analysis. In addition, the latest types of datasets used are not only technical and fundamental indicators, but also news, social media, macroeconomics including their combination. More importantly, the types of improvements and modifications that have been proposed in the study, are very important for developing predictions and filling research gaps.

So that it gives a complete contribution in systematic literature review papers including, the most significant journals, the most active and qualified researchers, research topics, types of data collections, types of methods, types of methods that are often used, types of best performing methods, types of method improvements and modifications, and types of methods. framework that is the purpose of this paper. The following is the arrangement of writing in this paper. In part 2, an introduction which contains the background and importance of the research is explained, the research methodology is explained in part 2, in part 3 presents the results and answers to the research questions. Section 4 or the last section is finally the conclusion of this paper in summary and the potential future work is explained.

2. Methodology

Review Method

The literature related to stock predictions is selected with a systematic approach. Within the scope of information science, systematic literature review is an important method and must be carried out. Collect all research evidence, identify, and assess and aim to answer the specific research question defined SLR [21]. This literature review

was conducted as a systematic and comprehensive literature review based on the guidelines proposed by Kitchenham [21]. The review methodology, stages, and some of the illustrations in this section are also inspired by the work [22], [23].

The three main phases of SLR preparation are illustrated in figure 1, they are: review planning, implementation, and reporting of results. The first step is to identify needs and requirements for a comprehensive review (step 1). In the introduction the purpose of the literature review has been described. Furthermore, the identification and review of the existing systematic reviews of stock predictions is carried out. It is possible that researcher bias could exist, so to guide the conduct of the review and reduce bias it is necessary to design a protocol (step 2). In this step, research questions are identified, paper search strategies, determination of inclusion and exclusion criteria for the study selection process, quality assessment, data extraction process from the main study, lastly perform data synthesis. During the implementation, assess and reporting stages of the review, protocol development review, evaluation, and iteratively be increased.

Research Questions

To keep the research focused, it is necessary to determine the research question. Adopting the PICOC approach [21] which stands for population, intervention, comparison, outcome, and context is designed to build the structure of the research question. Table 1 explanation the structure the PICOC questions of research (RQ).

Table 1. PICOC Structure Research Question

Structurer	Description
Population	Stock, stock time series, stock market, stock exchange, trading system
Intervention	Stock prediction, stock forecasting, technical analysis, fundamental analysis, classification, estimation, models, methods, techniques, datasets
Comparison	n/a
Outcomes	Prediction accuracy of stock return, successful stock prediction methods
Context	Studies in stock forecast, stock market decision, datasets type stock predict

From the main study, to answer RQ4 to RQ9 we extracted stock prediction algorithms, method development, dataset types, and prediction frameworks. Furthermore, the extraction results are analyzed to determination which ones are included and which are not related to the significance in stock market predictions.

Table 2. Motivations and research questions discussed in the literature review

Code	Research Questions	Motivation
RQ 1	Which journals with the highest and most significant number are the purpose of stock prediction studies?	Consider the journal with the most significant number in the main reference.
RQ 2	Who have contributed the most in the field of stock prediction methods?	Identify the researchers who have contributed the most to the field of stock market prediction research.
RQ 3	What are the trending research topics studied in the field of stock market prediction?	Trend analysis and research topics that emerged from stock prediction research.
RQ 4	What kind of datasets and features by which amount to predict stock trading?	Identify the types of data sets and data combinations that have a high probability of accuracy for stock trading predictions.
RQ 5	What are the different methods for solving stock market prediction challenges?	Identify research gaps related to stock prediction methods metode.
RQ 6	What type of method is most frequently adopted by researchers for stock prediction?	Identify the development potential and high performance of frequently used methods.
RQ 7	What are the current high-performing methods for stock prediction?	Identify new methods as potential developments and combinations of methods for better results.
RQ 8	What improvements and modifications have been made to improve stock prediction performance?	Identification of research gaps and differences that could potentially improve predictive performance.
RQ 9	For stock prediction, what kind of framework is available in the paper proposed by the researcher?	The available framework is proposed to identify gaps from one researcher to another.

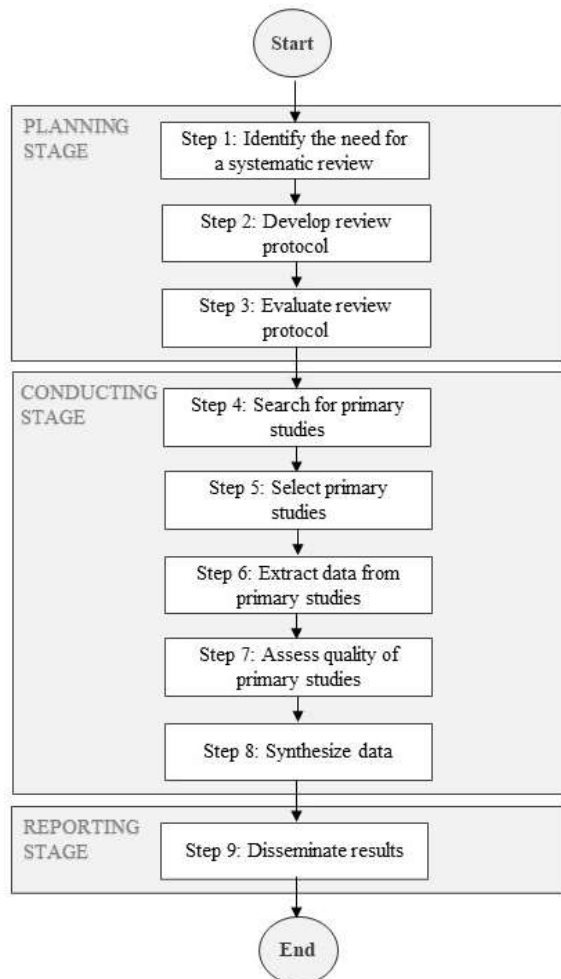


Fig. 1. Steps for a Systematic Literature Review

Table 2 shows the motivations and research questions discussed in the literature review. To help evaluate the context of the main study, the main research questions like in RQ4 to RQ9 and general questions lie in RQ1 to RQ3. RQ1 to

RQ3 provides us with summaries and synopsis of specific research areas in the field of stock exchange prediction.

The basic mind map was compiled to make it easier to see a comprehensive picture of the study shown in Figure 2 below. The types of data sets, frameworks, learning algorithms, and method improvements are identified which are the main objectives of this paper.



Fig. 2. Illustration of a basic mind map of a systematic literature review on stock prediction.

Search Strategy

Several processes in the search for papers (step 4) were carried out, including selecting a digital library, the search string was determined, running a search based on the string, updating the search string, and then some initial lists of studies were retrieved from the digital library that matched the search criteria.

Before the search begins, to increase the chances of finding suitable papers several screening sets should be set. The most popular digital library database in the scope of scientific knowledge is searched for various studies in the

world. A broad perspective is required for a broad coverage and extensive literature. The following is a list of digital databases that were searched: (1) SpringerLink (*springerlink.com*), (2) ScienceDirect (*sciencedirect.com*), (3) IEEE eXplore (*ieeexplore.ieee.org*), (4) Emerald (*emerald.com*), and (5) Taylor&Francis (*tandfonline.com*)

Using the following steps several strings were developed, (1) identification of the appropriate search space keywords from PICOC, focus terms from the intervention and population sections, (2) identify search terms from research questions, (3) search terms in titles and abstracts, and identified relevant keywords, (4) identifying search terms through synonyms, alternative vocabulary, and alternative spellings, (5) Construction of advanced search strings using identified search terms, Boolean AND and OR.

The following search string is finally used:

(stock OR shares OR exchange) AND (market OR price OR return) AND (forecast OR fundamental OR technical OR predict* OR probability OR assess* OR estimate* OR classificat*)*

Search string adjustments have been made, but the original is preserved, on the other hand engineering of the search string causes an increase in the frequency of irrelevant studies. Search in database by title, abstract and keywords. Searches are limited to the 2015 to 2020 publication period. The requirements of each database are tailored specifically to that search string. Only English language papers and journal papers are included, an exception to conference proceedings.

Study Selection

To select the main study inclusion and exclusion criteria were considered. Table 3 shows these criteria.

Mendeley Desktop is used to store and manage search results papers. Figure 3 shows each phase of the search process and the number of studies identified in detail. The two-step study selection process (step 5) is shown in Figure 3: (1) excluding the main study by title and abstract, (2) exclusion based on the full text of the paper. In addition, literature review studies and other studies that did not include experimental results were excluded. The degree of research similarity to the prediction of stock price movements, return benefits, and stock price forecasting was also included in the study.

Table 3. Inclusion and Exclusion Criteria

Inclusion Criteria	Studies in emerging markets and developed markets use all indicators (technical, fundamental, news, economy, etc.)
	A study that discusses and compares the performance of stock forecasting algorithms (machine learning, deep learning etc.)
	For studies of the type of research journal, excluding conference proceedings.
	With the same researchers and studies, only the most recent will be included. except, different topics and methods.
	The dataset used focuses on technical, economy, and fundamental analysis including momentum, if there is a combination of sentiment analysis is included.
Exclusion Criteria	Studies for which there are no performance measures for the machine learning field, and no validation.
	Studies that discuss stock forecasting data sets, methodology, improvement method in contexts other than stock predictions.
	Studies other than written in English.

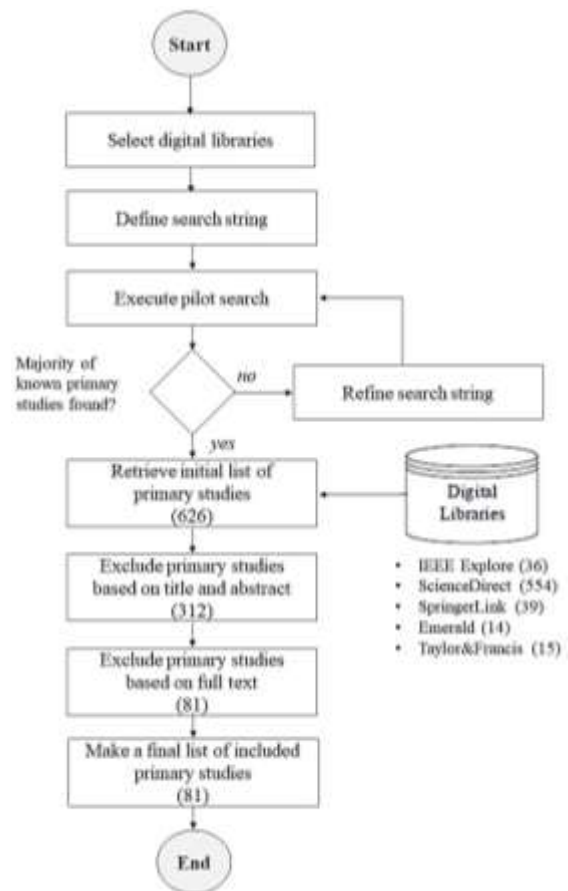


Fig. 3. Search and Selection of Main Studies

A total of 81 main study lists were generated from the first phase of the search process. Then, analyzes were carried out for the entire text in the papers of the 81 main studies. At the time of the analysis, the relevance to the research question and the quality of the research were taken into

account. At the end of this paper, an appendix with a complete list of the 81 selected studies (table 6) is presented.

Data Extraction

In collecting data that contribute to answering the research question, the main studies are extracted. To complete the data extraction, each of the 81 main studies was identified and analyzed into a special form designed to collect the study data needed to answer the research question (step 6). Table 4 shows a total of 6 properties identified through research questions and used to answer research questions, iteratively data extraction was carried out.

Table 4. Extraction Process Properties Data Mapped to RQ

Property	Research Questions
Publications & Researchers	RQ1, RQ2
Research Trends & Topics	RQ3
Stock Forecasting Datasets Type	RQ4
Stock Metrics	RQ4
Stock Prediction Algorithm,	RQ5, RQ6, RQ7, RQ8
Methods & Improvement Method	
Stock Prediction Frameworks	RQ9

Study Quality Assessment and Data Synthesis

To gather scientific evidence from studies that are screened to answer research questions is the goal of data synthesis. The process of assessing the quality of the study (step 8) to determine the strength of the conclusions outlined. Combining several scientific works of evidence can make an information strong, rather than one or two pieces of evidence that may be weak as information. In this review, the data extracted include qualitative data and quantitative data. Different strategies were used to synthesize the extracted data with regard to different types of research questions.

In general, the narrative synthesis method is used. Data is tabulated in a way that fits the questions. Several types of visualizations including histograms, bar charts, pie charts, and tables are used to improve the presentation of research trend distributions, learning algorithms, improve market prediction methods and model performance.

Threats to Validity

This SLR aims to analyze studies related to stock predictions based on machine learning algorithms, method improvements, types of stock prediction data sets, and frameworks. Bias in the study may exist, but it is not the scope of this study. The selection of published studies is not based on the filtering of all papers published in the journal but based on the related research

journals above. In practice, some conference or journal proceedings may not be detected if the study is outside the scope of this research, for example stock prediction studies in finance & management journals or accounting journals.

Studies from conference proceedings are not excluded from this SLR, as most studies are generally published in conference proceedings. Due to the increased workload significantly when reviewing studies some of the SLR papers did not use the conference proceedings as in [24]. In contrast to [18], [19] included conference proceedings as the main study in their systematic literature review.

3. Research Result

Significant Published Journal

A total of 81 major studies analyzing stock predictive performance are featured in this literature review. Figure 4 shows a real picture of the distribution of studies over the last five years from various digital libraries in the world, which is shown to show the level of interest of scientific researchers in the field of data science, stock prediction has increased from the past to the latest years. There is a significant increase in studies published in 2019, showing that there is a drastic trend of increasing interest. Figure 4 also shows that 2020 is still in June when this article was written, the number of publications is higher than in previous years meaning that the field of research on stock market predictions is still very relevant today and is getting a strong concern.

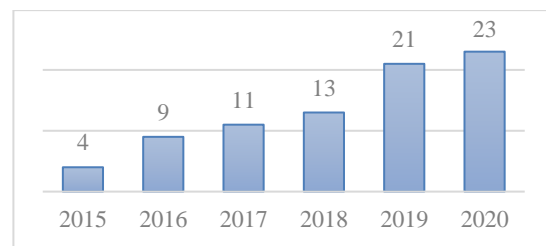


Fig. 4. Distribution of the main studies of the last five years

The international journals that are the target of publication are shown in Figure 5, based on the selected main studies. In this figure conference proceedings are excluded.

The graph above shows that the Expert Systems with Applications journal is at the top because it makes up the majority of stock prediction research papers in publications. Expert Systems With Applications is the best journal of international reputation, which publishes expert

systems and intelligent algorithms in various research groups in the world such as academic research in universities, industry and business around the world. In the field of computer and information science, this journal is very comprehensive in publishing related to knowledge management, data science, data analytics, business analytical algorithms, machine learning, deep neural networks, big data analytics, data mining, text mining, genetic algorithms, and

heuristic optimization as well. published in this journal.

In addition to these fields, this journal also has a good track record in terms of managing editors and reviewers. This journal scores 1.49 in the Scimago Journal Rank (SJR) and occupies the Q1 category in Artificial Intelligence. Therefore, the journal Expert Systems With Applications is very prestigious as the main goal and choice for researchers to publish their work.

Table 5. Scimago Journal Rank (SJR) from Selected Journals

No	Journal Publications	SJR	Q Category	Freq.
1	Decision Support Systems	1.92	Q1 in Information Systems and Management	1
2	IEEE Transactions on Knowledge and Data Eng.	1.78	Q1 in Computational Theory Mathematic	1
3	Knowledge-Based Systems	1.75	Q1 in Artificial Intelligence	1
4	Expert Systems with Applications	1.49	Q1 in Artificial Intelligence	17
5	Applied Soft Computing	1.41	Q1 in Software	5
6	Information Processing and Management	1.19	Q1 in Computer Science Applications	1
7	Neurocomputing	1.18	Q1 in Artificial Intelligence	2
8	Neural Computation	1.13	Q1 in Art and Humanities	1
9	Plos One	1.02	Q1 in Multidisciplinary	2
10	IEEE Systems Journal	0.98	Q1 in Computer Science Applications	1
11	Big Data	0.97	Q1 in Computer Science	1
12	Journal of Big Data	0.93	Q1 in Computer Networks and Commun.	1
13	Intelligent Systems in Accounting and Management	0.90	Q1 in Business, Management Accounting	1
14	Neural Computing and Applications	0.80	Q1 in Software	6
15	IEEE Access	0.78	Q3 in Computer Science	9
16	International Journal of Machine Learning & Cybernetics	0.78	Q1 in Computer Vision Pattern Recognized	1
17	Applied Intelligence	0.73	Q2 in Artificial Intelligence	1
18	Soft Computing	0.71	Q2 in Computer Science	8
19	Journal of Computational Science	0.61	Q1 in Computer Science	2
20	Data & Knowledge Engineering	0.57	Q2 in Information Systems and Management	1
21	Journal of Ambient Intelligence and Computing	0.54	Q1 in Computer Science	3
22	Computational Intelligence and Neuroscience	0.53	Q1 in Computer Science	1
23	Complexity	0.51	Q1 in Computer Science	1
24	Cybernetics and Systems	0.43	Q2 in Information Systems	1
25	Heliyon	0.43	Q1 in Multidisciplinary	1
26	Applied Sciences	0.42	Q1 in Engineering	1
27	Algorithms	0.36	Q3 in Computational Mathematics	1
28	International Journal of Communication Systems	0.33	Q2 in Electrical and Electronic Engineering	1
29	Applied Artificial Intelligence	0.32	Q3 in Artificial Intelligence	1
30	Mathematical Problems in Engineering	0.28	Q2 in Engineering	2
31	Turkish Journal of Electrical Eng. & Computer Science	0.22	Q3 in Computer Science	1
32	Computational and Mathematical Organization Theory	0.22	Q3 in Computational Mathematics	1
33	Modelling and Simulation in Engineering	0.19	Q2 in Engineering	1
34	Smart Science	0.19	Q2 in Engineering	1
35	Advances in Intelligent Systems and Computing	0.18	Q3 in Computer Science	2
36	Journal of Information Technology Research	0.18	Q3 in Computer Science	1

The second most significant journal is IEEE Access, this journal maintains a special section highlighting specific topics of IEEE interest. IEEE Access is published by the Institute of Electrical and Electronics Engineers (IEEE) with a concentration area with specialization in the application of science, technology, engineering,

and mathematics [26]. There is no doubt that the quality of the IEEE Access journal has a score of 0.78 in the Q1 category in Computer Science.

Soft computing journals get the third highest number of publications for stock prediction research. This journal is very popular among informatics engineering and information science,

published scientific works include software engineering studies, methodologies, data science, data analytics, algorithms and optimization, and soft computing foundations. A total of 36 international reputable journals are presented quantitatively in Figure 5, and qualitatively displayed the ranking scores in Table 5.

With the summary and analysis of this journal, it is hoped that it can become a reference for further researchers in finding research gaps, innovations, and contributions to their research more easily and with quality. Based on the main study extracted, the Scimago Journal Rank and Quartile category (Q1-Q4 scale) are shown in table 5. To facilitate analysis, published journals of international repute are arranged based on SJR scores from the highest to the lowest. For further research, focus on the top five journals for the search for scientific papers on stock predictions because there may still be many papers published in those journals in June 2020, the limit of the search for this paper.

The Most Contributing Researchers

It is necessary to investigate the researchers who made the most influential contributions to stock prediction research. The goal is that the researcher will become a role model and follow his scientific works. These researchers are most contributing and most active researchers are shown in Figure 6 below. The first and non-first author were selected based on the order in which they were written on the paper. All researchers were enrolled according to the main study.

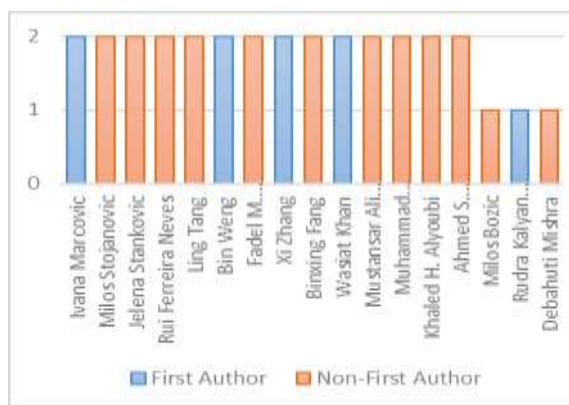


Fig. 5. The most contributing, influential, and number of research researchers for stock prediction

Research Topics in the Field of Stock Market Predictions

Stock market prediction is a significant research topic in the field of data mining and has

begun to place great emphasis on machine learning techniques as they exhibit a broad ability to simulate more complex problems [3]. Recent stock predictions focus on five topics, these are the findings and disclosures of a comprehensive analysis of the main studies.

Estimating stock price movements and returns in trading time series, using estimation algorithms or regression forecasting algorithms (Estimation/ Forecasting/Regression).

Finding the relationship between the emergence of bullish and bearish signal indicators for stock price movements using the Association rules algorithm.

Using a classification learning algorithm, classifying stock price movements usually into two to three classes such as “Up and Down”, or “Buy, Sell, and Hold” (classification).

Using a clustering algorithm, shares will be grouped against an investment decision making criterion (clustering).

Analyze and pre-process stock market data sets (Data Set Analysis).

Estimation is the first type of work, the approach used is regression or statistics, [10], Linear Regression [27], artificial neural networks [28] to estimate stock prices and profit returns in time series trading on the stock exchange. Estimation or forecasting results help as an important tool for contribute knowledge in the academic and financial environment [8], and can be used to support investor decision making in the selection of issuers that can provide short-term and long-term returns [29].

The second type of work (Association), uses the Associated Network algorithm to expose model associations in stock market predictions [30], this association method can be used to find relationships between signals that appear in technical indicators in stock predictions. which will be bullish.

The third type of work (Classification), classifies indicator data from a stock as “Buy”, “Hold”, or “Sell” through deep learning and neural network based classifications [31], [32], [33], [34], [16]. Prediction results can also represent an “Up” or “Down” trend so that investors can make decisions on investment entry positions by applying two single non-linear classifiers ANN, SVM and one RF ensemble approach to predict the direction of the next day's movement [35], [5], [36].

Clustering is the fourth type of work (clustering), using an unsupervised algorithm from machine learning that aims to group stocks based on certain input criteria, the K-Means algorithm proposed by Ying Xu, Cuijuan Yang, Shaoliang Peng, Yusuke Nojima (2020) to predict stock price movements [37]. Unsupervised learning methods such as grouping can be used to predict stock price movements, especially in improvising the merging of financial news data sets and social media to classify positive and negative sentiments.

The main focus of the fifth work (dataset analysis) is how to deal with data problems and pre-processing of data sets on indicators of stock price movements. Some researchers pre-process the datasets using several methods, while others analyze the stock indicator datasets in various aspects of viewpoint. [38] demonstrates and explains why dozens of technical indicator datasets require significant pre-processing to improve predictive accuracy. with the construction of the random forest method, they combined it with the method of treatment the problem of unbalanced data distribution. ntil now, it is still an NP-Hard problem to solve the superior combination of features [38], they use the forward sequential method to select candidate features.

The distribution of stock prediction research topics from January 2015 to June 2020 is shown in Figure 7. 56.79% of research studies are on estimation topics, 35.80% of studies focus on classification techniques, and 4.94% of main studies are related to analytical topics. datasets. Grouping and association is a minority research topic only 1.23%.

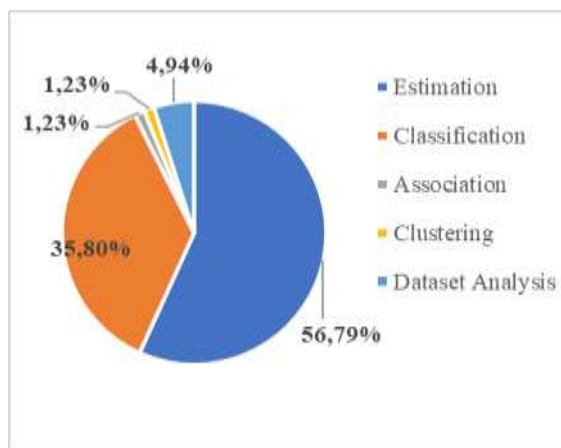


Fig. 6. Distribution of Research Topics

Finally, the topic of classification and estimation has great interest in stock prediction and in the future it is possible to improve the performance of that topic. hy researchers pay so much attention to this topic, here are three

possibilities: (1) the topic of estimation and classification is important according to the needs of the industry which requires several algorithm for predict which indicator signals very possible for give better returns. (2) related to technical datasets of indicators and fundamental indicators, most of which are ready to be used for classification and estimation methods. The third reason, (3) The performance of the clustering and association methods is less satisfactory than the stock price estimation and classification methods, therefore the clustering and association methods are very minimally published.

Datasets Type Used For Stock Predictions

For learning development purposes, certain data sets are used [39]. Different types of data sets provide different performance. More importantly, the treatment of the data set is very sensitive to the numerical results of the model measurements. Therefore, a collection of data that represents the object of research and its treatment before entering the modeling stage is very important and must be considered.

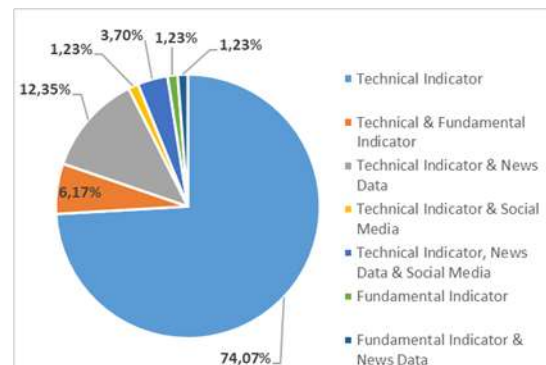


Fig. 7. Total Distribution of the Dataset

In this subsection, based on 81 main studies, the distribution of data set types from January 2015 to June 2020 is shown in Figure 8. A total of 74.07% of research studies use technical indicator datasets, being the largest percentage of the 7 dataset combinations that we found. Furthermore, 12.35% of research studies use a combination of technical indicators and news. These data sets are mostly located in public repositories of stock exchanges such as China, Korea, German, Netherlands, Great Britain, Japan, Hongkong, Canada, France, India, America Stock Exchange, etc. and are freely distributed.

Methods Used in Stock Market Predictions

Since 2015, 48 methods have been adopted and proposed as the best algorithm for stock

prediction as shown in Figure 9. The method is an application of five main topics of stock prediction.

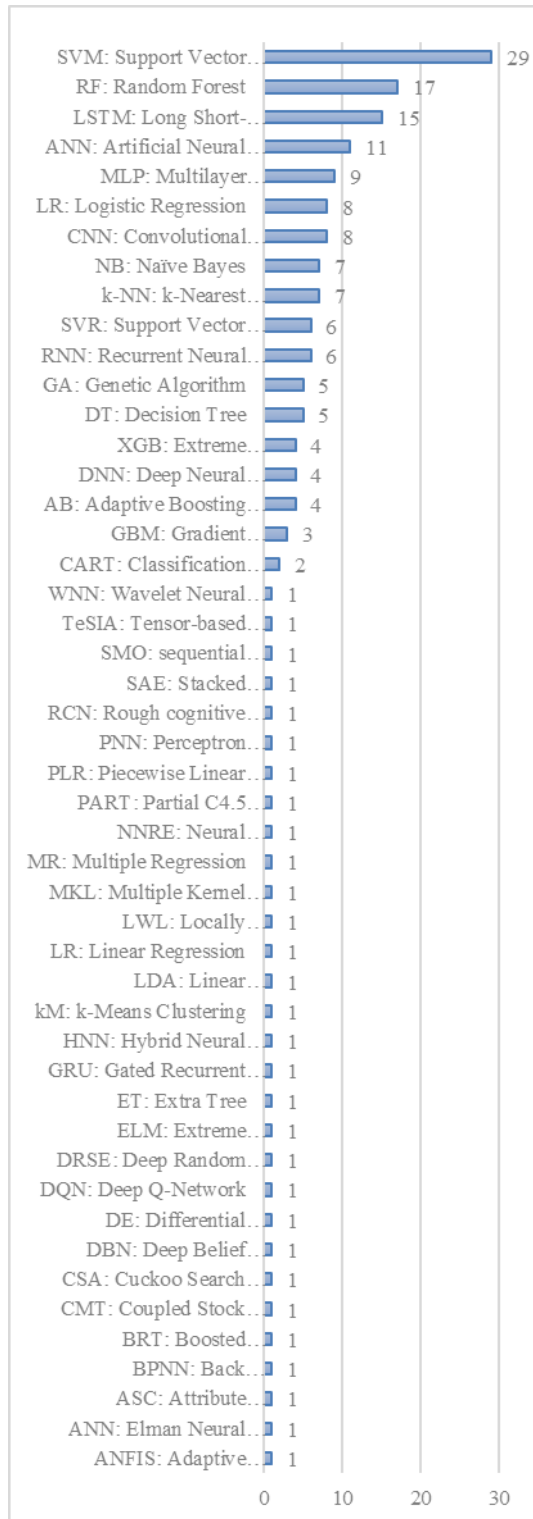


Fig. 8. Methods Used in Stock Exchange Predictions

The Most Often Used Method in Stock Prediction

Identification was carried out on 9 classification methods and the most widely adopted estimation of the 48 methods. Such methods include, ANN, LSTM, CNN, MLP, LR, RF, SVM, k-NN, and NB.

The percentages of SVM, RF, LSTM, and ANN are used by 65% of stock prediction research, meaning that this method still has the potential to be applied to stock prediction topics, will continue to be developed and receive special attention from researcher.

Best Performing Method for Stock Prediction

While many studies on predicting stock price movements individually report on measuring the performance of the proposed modeling methods, there is no strong consensus as to which method is best when each is looked at individually. Lee et al. [40] and Li et al. [41] concluded performance the SVM performs very well compared to other machine learning algorithms such as DT and Neural Network-based (MLP & LSTMNN) by producing a higher level of accuracy and return. high. However, Hall et al. highlights that studies using SVM underperformed. These may perform below expectations because they require parameter optimization for best performance [42].

RF seems to be the best performing method used in the field of stock prediction [4]. Likewise with Khan et al. [5] the random forest method provides high accuracy with a combination of types of technical data set indicators and related company news sentiment.

Several studies on stock market predictions show that as a predictor, the RMSE Neural Network value is very good [7], [43] in regression problems. There is also, to find the best hyper-parameter Neural Network, Genetic Algorithm is adopted. NN has proven to be more adequate to deal with complex and non-linear relationships between Stock Exchange metrics and stock indicators that fluctuate over time series [33]. On the other hand, it is still a chore to handle the appropriate parameters of the network architecture on the Neural Network, including learning rate, number of hidden neurons, momentum, and training cycles [44].

Finally, it can be concluded that the best performance results are obtained from the right method and the right pre-processing for the right data sets [4]. For all data, no specialist method is high-performing. Table 6 shows the methods, types of data sets, and performance results of each method that performed well.

Table 6. Top Performing Methods for Stock Market Predictions

Author	Method	Stock Market	Dataset Type	Accuracy	RMSE
[38]	RF (Ensemble Bagging-DT)	SZSE China	Technical	65,20%	-
[36]	XGB (Ensemble Boosting-DT)	S&P 500 USA	Technical	54,10%	-
[4]	RF (Ensemble Bagging-DT)	Nasdaq USA	Technical & News	67,00%	-
[45]	DNN	FTSE 100 UK	Technical	-	0,1160
[43]	CNN	CSI 100 China	Technical	-	0,1561
[11]	Stacking Meta-LR, Base-(RF, ERT, LGBM, RNN, BRNN, LSTM, GRU)	Nasdaq, Dow Jones, S&P 500	Technical & Macroeconomic	70,70%	-
[40]	SVM	S&P 500 USA	Technical	71,50%	-
[33]	CNN & Genetic Algorithm	Kospi Korea	Technical	73,74%	-
[5]	RF (Ensemble Bagging-DT)	S&P 500 USA	Technical & News	73,71%	-
[43]	CNN	CSI 100 China	Technical	-	0,1413
[46]	DNN & Harmony Search	SZSE China	Technical	74,10%	-
[47]	RNN	SZSE China	Technical	73,42%	-

Proposed Method Improvements and Modification for Stock Predictions

Researchers propose several techniques to improve the accuracy of machine learning classifiers for stock market predictions. To improve the prediction performance, several techniques are proposed by stock prediction researcher. Combining techniques (ensemble) using several learning methods [46], [47], [48], [49], [31], [11], [4], [5]. Using boost algorithm (Boosting Method) [50], [51], [13], [34], [52]. Added a feature selection method [53], [54], [55], [36], [56]. By using hyper-parameter optimization (tuning method of hyper-parameter) for several learners [57], [58], [10], [59], [60], [34], [61], [5].

However, although various stock prediction methods and improvement techniques have been proposed, none have been shown to perform consistently when predicting. consistently high-performing predictive methods are a major challenge for economists, investment managers, and data scientists. There is a crucial need for a consistently high-performing stock prediction framework that must be more robust against class imbalance, noise, and issues associated datasets.

Ensemble Machine Learning

Training multiple learning machines, combining their outputs, and making a final predictive decision from the combined outputs through weighting, majority voting, or using a meta-learner algorithm is an understanding of ensemble learning, treating multiple learning machines like decision-making "committees". [39]. Several empirical and theoretical studies prove that models built using the ensemble method achieve higher accuracy than single models. In research [11], one type of ensemble method was able to outperform the performance of the latest state of the art decision tree based and deep learning methods. Then the decisions of each ensemble member can be combined with several methods such as majority vote, average,

probabilistic, and weighting. Most types of ensemble methods can be applied very well in various types of learning tasks.

The Bagging and Boosting ensemble algorithm is among the effective and popular ensemble methods in machine learning. There is one more type of ensemble, namely Stacking with the Meta-Learner concept. Bagging and Boosting combine a single algorithm of the same type, while Stacking can combine different types of models. Stacking will perform well with several different models combined. Because when one model gives poor results at a point, but there are other models that get good results at that point, the combination of Stacking models will correct errors and improve performance results. In several studies, Meta-Learner built with the Stacking method has the ability to surpass several Decision Tree-Based and Deep Learning ensemble models.

Feature Selection

In order to improve the performance of machine learning techniques to reduce the dimension of the data is proposed, this is the study of feature selection. If a data set with f features and d dimensions, feature selection aims to reduce d to d' and $d' < d$ [39]. To reduce this dimension is the most frequently used approach. There is also a feature extraction approach that is quite effective, of course experiments need to be carried out to obtain the best results. For the feature extraction simulation, assuming it has six features $f1, f2, f3, f4, f5,$ and $f6$ if it produces 3 features, then the 3 selected features are a subset of the original 6 features (for example $f1, f3, f5$), but 3 extracted is a combination of 6 original features.

Some of the advantages of selecting features include improving the performance of learning algorithms, accelerating computing, and effective data collection. Only retain relevant features and eliminate irrelevant features which are commonly called noise without affecting learning performance, redundant features also include irrelevant ones.

Filters, wrappers and embedded are three popular feature selection methods. The filter method evaluates learners independently, ranks features based on statistical measure scores, and selects the best. Wrapper is known as a very expensive method, because to determine whether a feature should be selected or removed requires an evaluation by learning. The embedding method is like in decision tree induction where features must be selected first at each branching point. The filter and wrapper methods are most widely used when feature selection is carried out during data pre-processing.

In addition, to find out the best combination of features in terms of reducing dimensions, the following techniques can be used, (1) such as Subset Selection (feature ablation), this technique is a time-consuming brute force method, namely by trying all combinations of features, (2) Principal Components Analysis (PCA) or dimensionality reduction improves on previous techniques that focus on eliminating features, which can lead to the removal of potential features in the model not being able to understand the complexity of the problem. PCA transforms data from one space to another, where data is represented with smaller dimensions. With a note, data with smaller dimensions must be able to represent the characteristics of the data in the original dimensions, so that one feature in the new dimension may contain information about several features in the original dimension.

Parameter and Hyper-parameter Optimization

In machine learning, the presence of certain variables makes a learning algorithm much more flexible than the "Seeming Stupid Model", in the sense that it is able to map more data from the input domain to the output domain correctly. These variables are called parameters, weights, or coefficients. There is also a model component, namely hyper-parameters that affect the model output. Parameters and hyper-parameters that are set at random may not match the training data or maximum output, so optimization or tuning is necessary.

Several studies have proven that optimized parameters and hyper-parameters can improve model performance than models without optimization. Research [36] in hyper-parameter XGBoost algorithm optimized using Genetic Algorithm, [51] LSTM neural network parameters optimized using Random Search, [5] hyper-parameter Random Forest algorithm optimized using the Grid Search method, [52] experimented with four algorithms namely ANN, SVM, RF, and LR which were optimized using the Grid Search

method, as well as [43] the Deep Neural Network that was built optimized using the grid Search method.

Proposed Frameworks for Stock Prediction

A total of ten frameworks that have innovations accompanied by complete data mining stages in the last five years in the field of stock market prediction are described in the following sub-chapters.

Markovic et al.'s Framework

The results of their research [53] reveal that the Support vector Machine classifier has an average predictive performance of 60 percent for the three stock exchange markets (BELEX15, S&P500, FTSE100), after selecting attributes using the AHP method. The feature weighting results generated by the Analytical Hierarchy Process method are used for ranking and feature selection, and used with LS-SVM through a weighted kernel. The results of the performance measurement show that the proposed method performs better than the ANN and RF methods.

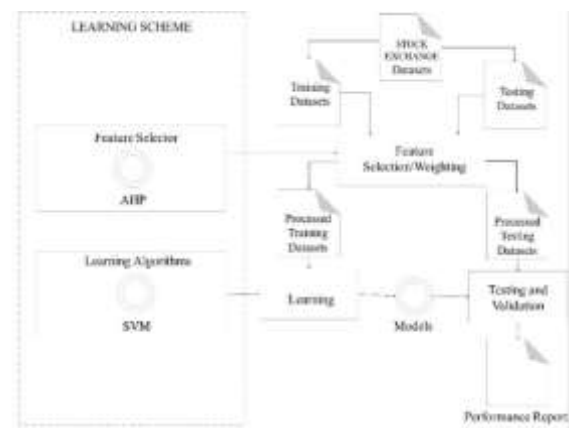


Fig. 9. Illustration of Markovic et al.'s framework

Kumar et al.'s Framework

Kumar also did the same with Markovic et al's framework [54]. However, Kumar et al. improve the selection of attributes when building a prediction model. Lessmann et al. consider four feature selection techniques: (1) Linear Correlation (LC), (2) Rank Correlation (RC), (3) Regression Relief (RR), (4) Random Forest (RF). The performance of the model is evaluated for twelve different stock indices. Figure 13 shows the prediction framework of Kumar et al.

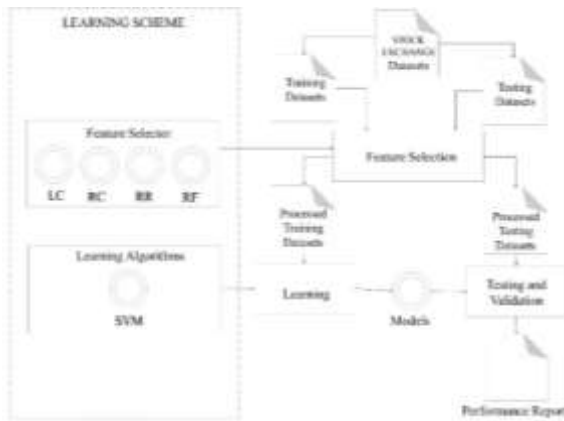


Fig. 10. Illustration of Kumar et al.'s Framework

Weng et al.'s Framework

Three stock prediction methods are compared in this study [55]. Their research reveals that the Support Vector Machine classifier has an average prediction accuracy of 82 percent, an average precision of 80 percent, after attribute selection using the Recursive Feature Elimination (RFE) algorithm. SVM-RFE significantly outperforms Neural Network and Decision Tree prediction methods. Figure 15 shows the predictive framework for Weng et al.

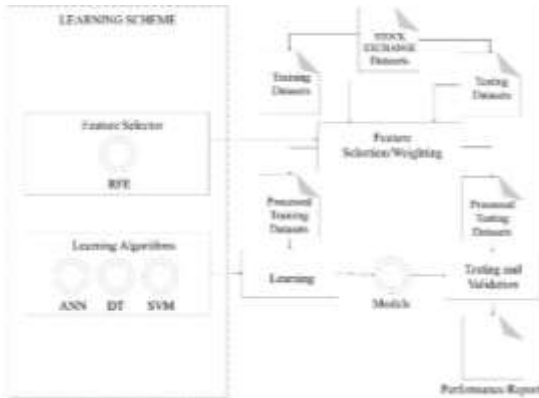


Fig. 11. Illustration of Weng et al.'s Framework

Chen et al.'s Framework

Conducting schematic evaluation and stock prediction are two special objectives of the development of the stock prediction framework by Chen et al., schematic evaluation focuses on evaluating the performance of learning algorithms, while stock prediction focuses on making final predictors using historical time series data according to the learning scheme. Then, new stock data is predicted using these predictors. The learning framework consists of: (1) pre-processed data, (2) attribute selection & dimension reduction, and (3) machine learning

algorithms. The main differences between Chen et al.'s frameworks, and the framework of Markovic et al. rests on the initial data preprocessor to create unique features, and feature selector selection using info gain. Cheng et al. also uses the additional k-Nearest Neighbor algorithm to improve the quality of predictions.

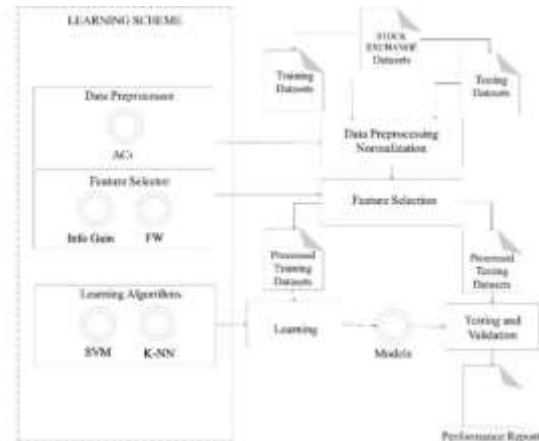


Fig. 12. Illustration of Chen et al.'s Framework

Weng et al.'s Framework

Weng et al., [13] also follow up on research results [53] and also the results of their previous research [55] about stock price predictions. In its development, Weng et al. In more detail, the preprocessor data includes cleaning data for existing missing and outliers and transformation data. The main differences between the frameworks of Weng et al. and the previous framework lies in the selection of attributes using a PCA selector by comparing the performance of four classification ensemble algorithm techniques to predict stock price movements.

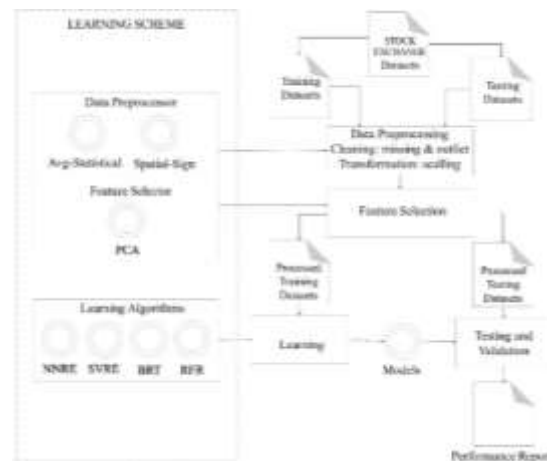


Fig. 13. Illustration of Weng et al.'s Framework

Guo et al.'s Framework

Different from other studies, [10] trying to use Particle Swarm Optimization (PSO) in selecting features. Furthermore, comparing the performance of two classification algorithm techniques to predict stock prices.

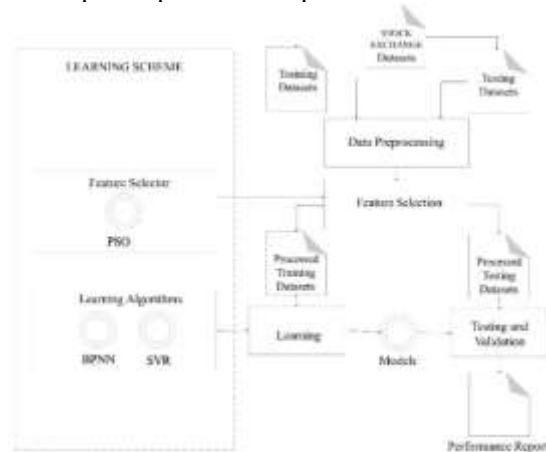


Fig. 14. Illustration of Guo et al.'s Framework

Zhang et al.'s Framework

Zhang et al. [38] published a study that focuses on learning stock price prediction schemes consisting of two section: scheme validation and stock price forecasting. The learning scheme is evaluated for its performance for schematic evaluation, while stock price forecasting focuses on making the final forecast using a time series data set according to the methodology. then predictors are used to forecast historical patterns of new stock price trends.

Learning method is composed of: (1) preprocessor data, (2) attribute selection, (3) balance class distribution, 4) learning algorithms. The four main differences from its predecessor framework lie in preprocessing data which uses pattern duration, model duration, and test duration. The feature selection in this study uses the forward sequential search (FSS) method, selecting only one of all candidates for all existing sections. [38] also improved the class imbalance problem in a classification, so they added the embeded RF algorithm method for the imbalance class problem distribution.

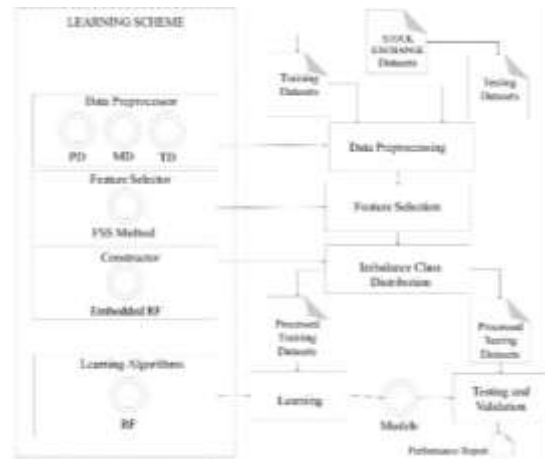


Fig. 15. Illustration of Zhang et al.'s Framework

Alsubaie et al.'s Framework

Research by Alsubaie et al. [57] compares the Naïve Bayes (NB) and Cost-sensitive Fine-tuned Naïve Bayes (CSFTNB) algorithm techniques to predict stock price movements. In feature selection, they compared Relief-F as the proposed method with Correlation and Gain Ratio.

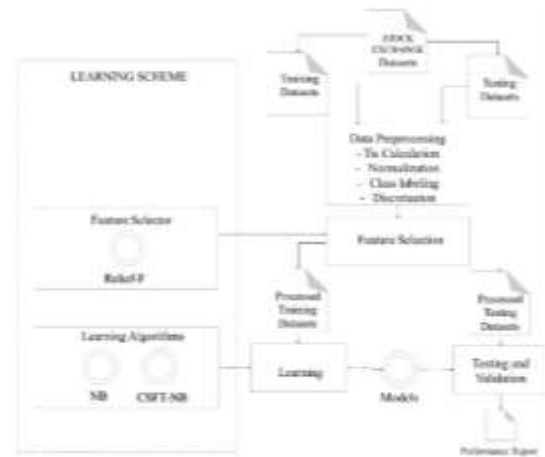


Fig. 16. Illustration of Alsubaie et al.'s Framework

Rustam et al.'s Framework

Rustam et al. [58] perform the attribute selection method using PSO as has been done by Guo et al. [10], after normalization of technical data indicators and support vector regression (SVR) learning algorithms. Feature selection using Particle Swarm Optimization showed superior performance against the experimental results of the study. The average RMSE measurement results are very small, which is below 0.1.

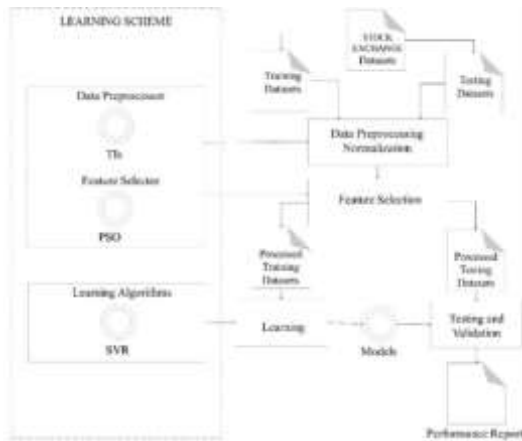


Fig. 17. Illustration of Rustam et al.'s Framework

Gong et al.'s Framework

Gong et al. [59] published a study comparing the performance of three prediction algorithm methods for predicting stock prices. Uses American (S&P 500) and Hong Kong (HS300) stock trading time series repositories. Their research reveals that the LS-SVM classifier has relatively small MAE, MSE, RMSE values, namely 0.00083, 0.0000029 and 0.00328 after selecting the PSO attribute.

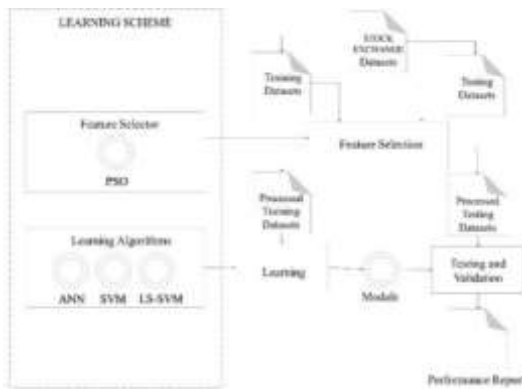


Fig. 18. Illustration of Gong et al.'s Framework

4. Conclusion and Future Work

This literature review identifies and analyzes research topic trends, types of data sets, learning algorithm, methods improvements, types of best performing methods, and frameworks used in stock exchange prediction. A total of 81 studies were investigated, which were published regarding stock predictions in the period January 2015 to June 2020 which took into account the inclusion and exclusion criteria.

Estimation or regression, clustering, association, classification, and preprocessing

analysis of data sets are the five main focuses revealed in the main study of stock prediction research. The classification method gets a share of 35.80% from related studies, the estimation method is 56.79%, data analytics is 4.94%, the rest is clustering and association is 1.23%. For this type of input data set, 74.07% of the research uses technical indicator datasets, then 12.35% of research studies use a combination of technical indicators and news data, the third place is 6.17% using a combination of technical indicators and fundamentals, the rest is below 5%. Therefore, there is still ample opportunity for further research on topics of high interest such as estimation and classification, as well as the types of datasets that have recently been used such as the combination of technical with news, as well as with macroeconomics and fundamentals need to be followed up more deeply.

To develop a stock prediction model 48 different methods have been applied, 9 of the most widely applied methods were identified. They are ANN, LSTM, CNN, MLP, LR, RF, SVM, k-NN and NB.

There is no strong consensus as to which method is best when each is looked at individually. The SVM using a technical dataset type indicator in [40] and [41] has reported that it works very well by producing an accuracy rate of 71.5% and a return of 13.9 % per week compared to RF with an accuracy rate of 66.2% and a return of 10.8% per week and LR with an accuracy rate of 65.8% and a return of 11.1%. RF is the best performing method in [3] combining types of technical data sets, indicators and company fundamentals. Likewise with Picasso et al. [4] combining the types of technical data set indicators and financial news (sentiment) RF method provides an accuracy rate of 67.0% and an annual return of 85.2% which is the same as work [5] level 73.71% accuracy for the combination of technical indicators and news in the RF method. The Neural Network research [46] using a type of technical data set indicator provides an accuracy rate of up to 74.1%, as well as the RNN in [47] shows an accuracy rate of 73.42 %. In another study, the ensemble Stacking technique was used, combining technical and macroeconomic indicators [11] using the four basic DecisionTree-Based algorithms with the Stacking meta-classifier LR method, the results showed RF 66.57%, ERT 70.07% , XGBoost 65.96%, LighGBM 66.87%, and Stacking-LR obtained 70.74% results, the change from base-classifier to meta-classifier was 0.67%.

Finally, it can be concluded that the best performance results are obtained from the right method and the right pre-processing for the right

data sets. For all data, no specialist method is high-performing. The results of this study also identify ten frameworks that are highly systematic and therefore influential in the field of prediction. They are Markovic et al. Framework, Kumar et al.'s Framework, Weng et al.'s Framework, Cheng et al.'s Framework, Guo et al.'s Framework, Zhang et al.'s Framework, Alsubaie et al.'s Framework, Rustam et al.'s Framework, and Gong et al.'s Framework. This framework can be considered for use and modification in future research.

In the future, from this systematic literature review we find and propose to perform a combination of input data set types for stock prediction because it has provided significant performance compared to a single data set type. In addition, it can contribute to combining several techniques to improve machine learning performance for stock market prediction; (1) by combining several machine learning methods through ensemble techniques, (2) by using boosting method, (3) using deep learning method because there are several developments of neural networks (4) by adding dimension reduction or feature engineering (selected) methods and (5) by using optimization or hyper-parameter tuning. We believe that there are still many research gaps that can be found for further research in improving stock prediction performance from three things, namely; (1) method selection, (2) data set type selection, and (3) machine learning performance enhancement modification techniques.

Finally, table 7 will present a list of 81 main studies from January 2015 to June 2020. The list consists of six columns including year, name of researcher, name of publication journal, type of data collection, method, and topic. To complete the basic mind map, it is necessary to present a complete mind map resulting from a systematic review of stock prediction literature.

Figure 22 is a complete mind map that has been used to comprehensively explore the relationship between the basic idea and the results of the exploration to answer the problem formulation. To see the big picture of all relevant issues and analyze options, a complete mind map form is a new perspective [60]. A complete mind map will make it easier to manage knowledge information comprehensively and integrate new scientific work. In this study, mind maps are used to present the results of a systematic literature review related to stock predictions.

Acknowledgement

I'm deeply indebted to my supervisor, Prof. Ir. Arif Djunaidy, M.Sc., Ph.D. Head of Laboratory of Data Engineer and Business Intelligence, ITS Surabaya for warm support, inspiration and thoughtful guidance. In addition to Dr. Ir. Dimitri Mahayana M.Eng through the CDSM™ (Certified Data Science for Manager) certification exam of PT Sharing Vision Indonesia has recognized me as a professional Data Scientist and Big Data Analytics.

References

- [1] C. Luo, C. Tan, X. Wang, and Y. Zheng, "An evolving recurrent interval type-2 intuitionistic fuzzy neural network for online learning and time series prediction," *Appl. Soft Comput. J.*, vol. 78, pp. 150–163, 2019, doi: 10.1016/j.asoc.2019.02.032.
- [2] C. Ji *et al.*, "A fast shapelet selection algorithm for time series classification," *Comput. Networks*, vol. 148, pp. 231–240, 2019, doi: 10.1016/j.comnet.2018.11.031.
- [3] Z. Tan, Z. Yan, and G. Zhu, "Stock selection with random forest: An exploitation of excess return in the Chinese stock market," *Heliyon*, vol. 5, no. 8, p. e02310, 2019, doi: 10.1016/j.heliyon.2019.e02310.
- [4] A. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical analysis and sentiment embeddings for market trend prediction," *Expert Syst. Appl.*, 2019, doi: 10.1016/j.eswa.2019.06.014.
- [5] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media, news," *J. Ambient Intell. Humaniz. Comput.*, 2020, doi: 10.1007/s12652-020-01839-w.
- [6] A. R. Azhikodan, A. G. K. Bhat, and M. V. Jadhav, "Stock trading bot using deep reinforcement learning," in *Lecture Notes in Networks and Systems*, 2019.
- [7] S. S. Yu, S. W. Chu, Y. K. Chan, and C. M. Wang, "Share Price Trend Prediction Using CRNN with LSTM Structure," *Smart Sci.*, vol. 7, no. 3, pp. 189–197, 2019, doi: 10.1080/23080477.2019.1605474.
- [8] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," *Neural Comput. Appl.*, 2019, doi: 10.1007/s00521-019-04212-x.
- [9] S. Kim, S. Ku, W. Chang, W. Chang, W. Chang, and J. W. Song, "Predicting the Direction of US Stock Prices Using Effective Transfer Entropy and Machine Learning Techniques," *IEEE Access*, vol. 8, pp. 111660–111682, 2020, doi: 10.1109/ACCESS.2020.3002174.
- [10] Y. Guo, S. Han, C. Shen, Y. Li, X. Yin, and Y. Bai, "An adaptive SVR for high-frequency stock price forecasting," *IEEE Access*, vol. 6, pp. 11397–11404, 2018, doi: 10.1109/ACCESS.2018.2806180.
- [11] M. Jiang, J. Liu, L. Zhang, and C. Liu, "An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms," *Phys. A Stat. Mech. its Appl.*, 2020, doi: 10.1016/j.physa.2019.122272.
- [12] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, "Leveraging social media news to predict stock index movement using RNN-boost," *Data Knowl. Eng.*, 2018, doi: 10.1016/j.datak.2018.08.003.
- [13] B. Weng, L. Lu, X. Wang, F. M. Megahed, and W.

- Martinez, "Predicting short-term stock prices using ensemble methods and online data sources," *Expert Syst. Appl.*, 2018, doi: 10.1016/j.eswa.2018.06.016.
- [14] W. Khan, U. Malik, M. A. Ghazanfar, M. A. Azam, K. H. Alyoubi, and A. S. Alfakheh, "Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis," *Soft Comput.*, 2019, doi: 10.1007/s00500-019-04347-y.
- [15] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," *Inf. Process. Manag.*, 2020, doi: 10.1016/j.ipm.2020.102212.
- [16] X. Hou, K. Wang, J. Zhang, and Z. Wei, "An Enriched Time-Series Forecasting Framework for Long-Short Portfolio Strategy," *IEEE Access*, vol. 8, pp. 31992–32002, 2020, doi: 10.1109/ACCESS.2020.2973037.
- [17] X. Yuan, J. Yuan, T. Jiang, and Q. U. Ain, "Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market," *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.2969293.
- [18] O. Bustos and A. Pomares-Quimbaya, "Stock market movement forecast: A Systematic review," *Expert Systems with Applications*. 2020, doi: 10.1016/j.eswa.2020.113464.
- [19] D. P. Gandhmal and K. Kumar, "Systematic analysis and review of stock market prediction techniques," *Computer Science Review*. 2019, doi: 10.1016/j.cosrev.2019.08.001.
- [20] G. Kumar, S. Jain, and U. P. Singh, *Stock Market Forecasting Using Computational Intelligence: A Survey*, no. 0123456789. Springer Netherlands, 2020.
- [21] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering; Technical report EBSE-2007-01; EBSE: UK, Durham," 2007.
- [22] M. Unterkalmsteiner, T. Gorschek, A. K. M. M. Islam, C. K. Cheng, R. B. Permadi, and R. Feldt, "Evaluation and measurement of software process improvement-A systematic literature review," *IEEE Transactions on Software Engineering*. 2012, doi: 10.1109/TSE.2011.26.
- [23] D. Radjenović, M. Heričko, R. Torkar, and A. Živković, "Software fault prediction metrics: A systematic literature review," *Information and Software Technology*. 2013, doi: 10.1016/j.infsof.2013.02.009.
- [24] S. Teixeira Zavadzki, M. Kleina, F. Oscar Drozda, and M. Augusto Mendes Marques, "Computational Intelligence Techniques Used for Stock Market Prediction: A Systematic Review," *IEEE Latin America Transactions*. 2020, doi: 10.1109/TLA.2020.9082218.
- [25] F. Soleymani and E. Paquet, "Financial portfolio optimization with online deep reinforcement learning and restricted stacked autoencoder—DeepBreath," *Expert Syst. Appl.*, vol. 156, 2020, doi: 10.1016/j.eswa.2020.113456.
- [26] H. D. Chacon, E. Kesici, and P. Najafirad, "Improving Financial Time Series Prediction Accuracy Using Ensemble Empirical Mode Decomposition and Recurrent Neural Networks," *IEEE Access*, vol. 8, pp. 117133–117145, 2020, doi: 10.1109/ACCESS.2020.2996981.
- [27] I. R. Parry, S. S. Khurana, M. Kumar, and A. A. Altalbe, "Time series data analysis of stock price movement using machine learning techniques," *Soft Comput.*, vol. 0123456789, 2020, doi: 10.1007/s00500-020-04957-x.
- [28] M. Nikou, G. Mansourfar, and J. Bagherzadeh, "Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms," *Intell. Syst. Accounting, Financ. Manag.*, 2019, doi: 10.1002/isaf.1459.
- [29] F. D. Paiva, R. T. N. Cardoso, G. P. Hanaoka, and W. M. Duarte, "Decision-making for financial trading: A fusion approach of machine learning and portfolio selection," *Expert Syst. Appl.*, vol. 115, pp. 635–655, 2018, doi: 10.1016/j.eswa.2018.08.003.
- [30] G. Ding and L. Qin, "Study on the prediction of stock price based on the associated network model of LSTM," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 6, pp. 1307–1317, 2019, doi: 10.1007/s13042-019-01041-1.
- [31] H. Park, M. K. Sim, and D. G. Choi, "An intelligent financial portfolio trading strategy using deep Q-learning," *Expert Syst. Appl.*, vol. 158, p. 113573, 2020, doi: 10.1016/j.eswa.2020.113573.
- [32] P. A. Kanade, "Machine Learning Model for Stock Market Prediction," *Int. J. Res. Appl. Sci. Eng. Technol.*, 2020, doi: 10.22214/ijraset.2020.6030.
- [33] H. Chung and K. shik Shin, "Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction," *Neural Comput. Appl.*, 2019, doi: 10.1007/s00521-019-04236-3.
- [34] K. Sai Sravani and P. RajaRajeswari, "Prediction of stock market exchange using LSTM algorithm," *Int. J. Sci. Technol. Res.*, 2020.
- [35] P. Misra and S. Chaurasia, "Data-driven trend forecasting in stock market using machine learning techniques," *J. Inf. Technol. Res.*, 2020, doi: 10.4018/JITR.2020010109.
- [36] J. Nobre and R. F. Neves, "Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to trade in the financial markets," *Expert Syst. Appl.*, vol. 125, pp. 181–194, 2019, doi: 10.1016/j.eswa.2019.01.083.
- [37] Y. Xu, C. Yang, S. Peng, and Y. Nojima, "A hybrid two-stage financial stock forecasting algorithm based on clustering and ensemble learning," *Appl. Intell.*, 2020, doi: 10.1007/s10489-020-01766-5.
- [38] J. Zhang, S. Cui, Y. Xu, Q. Li, and T. Li, "A novel data-driven stock price trend prediction system," *Expert Syst. Appl.*, 2018, doi: 10.1016/j.eswa.2017.12.026.
- [39] G. I. Webb *et al.*, "Leave-One-Out Cross-Validation," in *Encyclopedia of Machine Learning*, 2011.
- [40] T. K. Lee, J. H. Cho, D. S. Kwon, and S. Y. Sohn, "Global stock market investment strategies based on financial network indicators using machine learning techniques," *Expert Syst. Appl.*, vol. 117, pp. 228–242, 2019, doi: 10.1016/j.eswa.2018.09.005.
- [41] Z. Li and V. Tam, "A machine learning view on momentum and reversal trading," *Algorithms*, 2018, doi: 10.3390/a11110170.
- [42] X. Zhang, Y. Li, S. Wang, B. Fang, and P. S. Yu, "Enhancing stock market prediction with extended coupled hidden Markov model over multi-sourced data," *Knowl. Inf. Syst.*, 2019, doi: 10.1007/s10115-018-1315-6.
- [43] Y. Ma, R. Han, and W. Wang, "Prediction-Based Portfolio Optimization Models Using Deep Neural Networks," *IEEE Access*, vol. 8, pp. 115393–115405, 2020, doi: 10.1109/ACCESS.2020.3003819.
- [44] S. Lessmann, B. Baesens, C. Mues, and S. Pietsch, "Benchmarking classification models for software defect prediction: A proposed framework and novel findings," *IEEE Trans. Softw. Eng.*, vol. 34, no. 4, pp. 485–496, 2008, doi: 10.1109/TSE.2008.35.
- [45] W. Wang, W. Li, N. Zhang, and K. Liu, "Portfolio formation with preselection using deep learning from

- long-term financial data,” *Expert Syst. Appl.*, vol. 143, p. 113042, 2019, doi: 10.1016/j.eswa.2019.113042.
- [46] X. Li and C. Luo, “An intelligent stock trading decision support system based on rough cognitive reasoning,” *Expert Syst. Appl.*, vol. 160, p. 113763, 2020, doi: 10.1016/j.eswa.2020.113763.
- [47] Y. Qiu, H. Y. Yang, S. Lu, and W. Chen, “A novel hybrid model based on recurrent neural networks for stock market timing,” *Soft Comput.*, 2020, doi: 10.1007/s00500-020-04862-3.
- [48] I. H. Witten, E. Frank, and M. a. Hall, *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*. 2011.
- [49] O. Maimon and L. Rokach, *Data mining and knowledge discovery handbook*, vol. 48, no. 10. 2011.
- [50] T. M. Khoshgoftaar and J. Van Hulse, “Empirical case studies in attribute noise detection,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, 2009, doi: 10.1109/TSMCC.2009.2013815.
- [51] Y. Wang, H. Liu, Q. Guo, S. Xie, and X. Zhang, “Stock volatility prediction by hybrid neural network,” *IEEE Access*, vol. 7, pp. 154524–154534, 2019, doi: 10.1109/ACCESS.2019.2949074.
- [52] A. Hasan, O. Kalipsiz, and S. Akyokus, “Modeling Traders’ Behavior with Deep Learning and Machine Learning Methods: Evidence from BIST 100 Index,” *Complexity*, 2020, doi: 10.1155/2020/8285149.
- [53] I. Marković, M. Stojanović, J. Stanković, and M. Stanković, “Stock market trend prediction using AHP and weighted kernel LS-SVM,” *Soft Comput.*, vol. 21, no. 18, pp. 5387–5398, 2016, doi: 10.1007/s00500-016-2123-0.
- [54] D. Kumar, S. S. Meghwani, and M. Thakur, “Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets,” *J. Comput. Sci.*, 2016, doi: 10.1016/j.jocs.2016.07.006.
- [55] B. Weng, M. A. Ahmed, and F. M. Megahed, “Stock market one-day ahead movement prediction using disparate data sources,” *Expert Syst. Appl.*, 2017, doi: 10.1016/j.eswa.2017.02.041.
- [56] Y. Chen and Y. Hao, “A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction,” *Expert Syst. Appl.*, 2017, doi: 10.1016/j.eswa.2017.02.044.
- [57] Y. Alsubaie, K. El Hindi, and H. Alsalman, “Cost-Sensitive Prediction of Stock Price Direction: Selection of Technical Indicators,” *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2945907.
- [58] Z. Rustam and P. Kintandani, “Application of Support Vector Regression in Indonesian Stock Price Prediction with Feature Selection Using Particle Swarm Optimisation,” *Model. Simul. Eng.*, 2019, doi: 10.1155/2019/8962717.
- [59] X. L. Gong, X. H. Liu, X. Xiong, and X. T. Zhuang, “Forecasting stock volatility process using improved least square support vector machine approach,” *Soft Comput.*, vol. 23, no. 22, pp. 11867–11881, 2019, doi: 10.1007/s00500-018-03743-0.
- [60] M. Manwal and A. Gupta, “Big data and hadoop-A technological survey,” 2018, doi: 10.1109/ICETCCT.2017.8280345.
- [61] I. Marković, M. Stojanović, M. Božić, and J. Stanković, “Stock Market Trend Prediction Based on the LS-SVM Model Update Algorithm,” 2015, doi: 10.1007/978-3-319-09879-1_11.
- [62] R. K. Nayak, D. Mishra, and A. K. Rath, “A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices,” *Appl. Soft Comput. J.*, 2015, doi: 10.1016/j.asoc.2015.06.040.
- [63] J. M. Pinto, R. F. Neves, and N. Horta, “Boosting Trading Strategies performance using VIX indicator together with a dual-objective Evolutionary Computation optimizer,” *Expert Syst. Appl.*, 2015, doi: 10.1016/j.eswa.2015.04.056.
- [64] M. Thenmozhi and G. Sarath Chand, “Forecasting stock returns based on information transmission across global markets using support vector machines,” *Neural Comput. Appl.*, 2015, doi: 10.1007/s00521-015-1897-9.
- [65] M. Göçken, M. Özçalici, A. Boru, and A. T. Dosdoğru, “Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction,” *Expert Syst. Appl.*, 2016, doi: 10.1016/j.eswa.2015.09.029.
- [66] Y. J. Chen, Y. M. Chen, and C. L. Lu, “Enhancement of stock market forecasting using an improved fundamental analysis-based approach,” *Soft Comput.*, vol. 21, no. 13, pp. 3735–3757, 2016, doi: 10.1007/s00500-016-2028-y.
- [67] M. Inthachot, V. Boonjing, and S. Intakosum, “Artificial Neural Network and Genetic Algorithm Hybrid Intelligence for Predicting Thai Stock Price Index Trend,” *Computational Intelligence and Neuroscience*. 2016, doi: 10.1155/2016/3045254.
- [68] M. Qiu and Y. Song, “Predicting the direction of stock market index movement using an optimized artificial neural network model,” *PLoS One*, 2016, doi: 10.1371/journal.pone.0155133.
- [69] C. H. Su and C. H. Cheng, “A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock,” *Neurocomputing*, 2016, doi: 10.1016/j.neucom.2016.03.068.
- [70] L. Yu, L. Hu, and L. Tang, “Stock Selection with a Novel Sigmoid-Based Mixed Discrete-Continuous Differential Evolution Algorithm,” *IEEE Trans. Knowl. Data Eng.*, 2016, doi: 10.1109/TKDE.2016.2545660.
- [71] X. Zhang, A. Li, and R. Pan, “Stock trend prediction based on a new status box method and AdaBoost probabilistic support vector machine,” *Appl. Soft Comput. J.*, 2016, doi: 10.1016/j.asoc.2016.08.026.
- [72] H. Gündüz, Z. Çataltepe, and Y. Yaslan, “Stock daily return prediction using expanded features and feature selection,” *Turkish J. Electr. Eng. Comput. Sci.*, 2017, doi: 10.3906/elk-1704-256.
- [73] E. Chong, C. Han, and F. C. Park, “Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies,” *Expert Syst. Appl.*, 2017, doi: 10.1016/j.eswa.2017.04.030.
- [74] J. J. Flores, J. R. C. González, R. L. Farias, and F. Calderon, “Evolving nearest neighbor time series forecasters,” *Soft Comput.*, vol. 23, no. 3, pp. 1039–1048, 2017, doi: 10.1007/s00500-017-2822-1.
- [75] R. Ghasemiyeh, R. Moghdani, and S. S. Sana, “A Hybrid Artificial Neural Network with Metaheuristic Algorithms for Predicting Stock Price,” *Cybern. Syst.*, vol. 48, no. 4, pp. 365–392, 2017, doi: 10.1080/01969722.2017.1285162.
- [76] L. Lei, “Wavelet Neural Network Prediction Method of Stock Price Trend Based on Rough Set Attribute Reduction,” *Appl. Soft Comput. J.*, 2017, doi: 10.1016/j.asoc.2017.09.029.
- [77] L. Luo, S. You, Y. Xu, and H. Peng, “Improving the integration of piece wise linear representation and weighted support vector machine for stock trading signal prediction,” *Appl. Soft Comput. J.*, 2017, doi: 10.1016/j.asoc.2017.03.007.
- [78] S. Pyo, J. Lee, M. Cha, and H. Jang, “Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets,” *PLoS One*, 2017, doi:

- 10.1371/journal.pone.0188107.
- [79] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, and P. S. Yu, "Improving stock market prediction via heterogeneous information fusion," *Knowledge-Based Syst.*, 2017, doi: 10.1016/j.knsys.2017.12.025.
- [80] X. Zhang, J. Shi, D. Wang, and B. Fang, "Exploiting investors social network for stock prediction in China's market," *J. Comput. Sci.*, 2017, doi: 10.1016/j.jocs.2017.10.013.
- [81] T. Gao and Y. Chai, "Improving stock closing price prediction using recurrent neural network and technical indicators," *Neural Computation*. 2018, doi: 10.1162/neco_a_01124.
- [82] G. H. Kim and S. H. Kim, "Variable Selection for Artificial Neural Networks with Applications for Stock Price Prediction," *Appl. Artif. Intell.*, vol. 33, no. 1, pp. 54–67, 2018, doi: 10.1080/08839514.2018.1525850.
- [83] P. P. S. Kohli, S. Zargar, S. Arora, and P. Gupta, "Stock prediction using machine learning algorithms," in *Advances in Intelligent Systems and Computing*, 2018.
- [84] K. H. Nam and N. Y. Seong, "Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market," *Decis. Support Syst.*, vol. 117, pp. 100–112, 2018, doi: 10.1016/j.dss.2018.11.004.
- [85] R. Ren, D. D. Wu, and D. D. Wu, "Forecasting stock market movement direction using sentiment analysis and support vector machine," *IEEE Syst. J.*, vol. 13, no. 1, pp. 760–770, 2018, doi: 10.1109/JSYST.2018.2794462.
- [86] Q. Wang, W. Xu, and H. Zheng, "Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles," *Neurocomputing*, 2018, doi: 10.1016/j.neucom.2018.02.095.
- [87] X. Zhou, Z. Pan, G. Hu, S. Tang, and C. Zhao, "Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets," *Math. Probl. Eng.*, 2018, doi: 10.1155/2018/4907423.
- [88] M. Z. Asghar, F. Rahman, F. M. Kundi, and S. Ahmad, "Development of stock market trend prediction system using multiple regression," *Comput. Math. Organ. Theory*, 2019, doi: 10.1007/s10588-019-09292-7.
- [89] J. Cao and J. Wang, "Stock price forecasting model based on modified convolution neural network and financial time series analysis," *Int. J. Commun. Syst.*, 2019, doi: 10.1002/dac.3987.
- [90] E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," *Expert Syst. Appl.*, 2019, doi: 10.1016/j.eswa.2019.03.029.
- [91] D. Lv, S. Yuan, M. Li, and Y. Xiang, "An Empirical Study of Machine Learning Algorithms for Stock Daily Trading Strategy," *Math. Probl. Eng.*, 2019, doi: 10.1155/2019/7816154.
- [92] A. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical analysis and sentiment embeddings for market trend prediction," *Expert Syst. Appl.*, vol. 135, pp. 60–70, 2019, doi: 10.1016/j.eswa.2019.06.014.
- [93] Q. Xu, L. Wang, C. Jiang, and Y. Liu, "A novel (U)MIDAS-SVR model with multi-source market sentiment for forecasting stock returns," *Neural Comput. Appl.*, vol. 32, no. 10, pp. 5875–5888, 2019, doi: 10.1007/s00521-019-04063-6.
- [94] F. Yang, Z. Chen, J. Li, and L. Tang, "A novel hybrid stock selection method with stock prediction," *Appl. Soft Comput. J.*, 2019, doi: 10.1016/j.asoc.2019.03.028.
- [95] C. Xiao, W. Xia, and J. Jiang, "Stock price forecast based on combined model of ARI-MA-LS-SVM," *Neural Comput. Appl.*, vol. 32, no. 10, pp. 5379–5388, 2019, doi: 10.1007/s00521-019-04698-5.
- [96] S. Kumar Chandar, "Grey Wolf optimization-Elman neural network model for stock price prediction," *Soft Comput.*, vol. 2, 2020, doi: 10.1007/s00500-020-05174-2.
- [97] F. Kamalov, "Forecasting significant stock price changes using neural networks," *Neural Comput. Appl.*, vol. 3, 2020, doi: 10.1007/s00521-020-04942-3.
- [98] M. Ananthi and K. Vijayakumar, "Stock market analysis using candlestick regression and market trend prediction (CKRM)," *J. Ambient Intell. Humaniz. Comput.*, no. 0123456789, 2020, doi: 10.1007/s12652-020-01892-5.
- [99] J. Zhao, D. Zeng, S. Liang, H. Kang, and Q. Liu, "Prediction model for stock price trend based on recurrent neural network," *J. Ambient Intell. Humaniz. Comput.*, 2020, doi: 10.1007/s12652-020-02057-0.
- [100] Q. Chen, W. Zhang, and Y. Lou, "Forecasting Stock Prices Using a Hybrid Deep Learning Model Integrating Attention Mechanism, Multi-Layer Perceptron, and Bidirectional Long-Short Term Memory Neural Network," *IEEE Access*, vol. 8, pp. 117365–117376, 2020, doi: 10.1109/ACCESS.2020.3004284.
- [101] M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, S. Member, and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," pp. 150199–150212, 2020, doi: 10.1109/ACCESS.2020.3015966.
- [102] Y. Hao and Q. Gao, "Predicting the trend of stock market index using the hybrid neural network based on multiple time scale feature learning," *Appl. Sci.*, 2020, doi: 10.3390/app10113961.
- [103] A. Kelotra and P. Pandey, "Stock Market Prediction Using Optimized Deep-ConvLSTM Model," *Big Data*, 2020, doi: 10.1089/big.2018.0143.
- [104] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A comprehensive evaluation of ensemble learning for stock-market prediction," *J. Big Data*, 2020, doi: 10.1186/s40537-020-00299-5.

Attachment 1

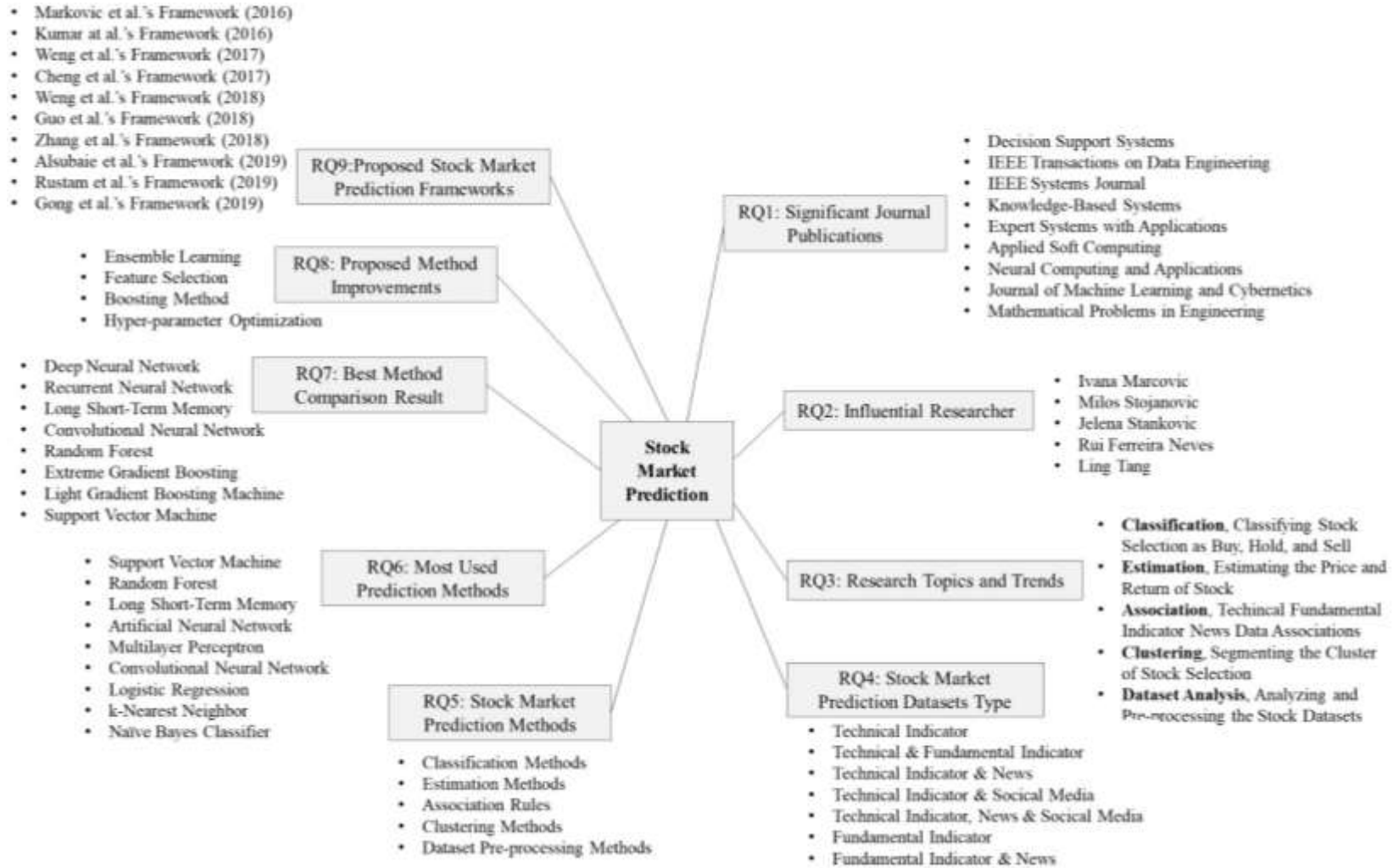


Fig. 19. Complete SLR Mind Map on Stock Market Predictions

Attachment 2

Table 7. List of Major Studies in Stock Market Prediction

Year	Primary Studies	Datasets Type	Methods	Topics
2015	[61]	Technical Indicator	SVM: Support Vector Machine	Classification
2015	[62]	Technical Indicator	SVM: Support Vector Machine k-NN: k-Nearest Neighbor	Estimation
2015	[63]	Technical & Fundamental Indicator	GA: Genetic Algorithm	Estimation
2015	[64]	Technical Indicator	SVM: Support Vector Machine	Estimation
2016	[53]	Technical Indicator	SVM: Support Vector Machine	Estimation
2016	[65]	Technical Indicator	ANN: Artificial Neural Network	Estimation
2016	[66]	Fundamental Indicator & News Data	NB: Naïve Bayesian Classifier	Classification
2016	[67]	Technical Indicator	ANN: Artificial Neural Network GA: Genetic Algorithm	Classification
2016	[54]	Technical Indicator	SVM: Support Vector Machine	Classification
2016	[68]	Technical Indicator	ANN: Artificial Neural Network GA: Genetic Algorithm	Estimation
2016	[69]	Technical Indicator	ANFIS: Adaptive Neuro Fuzzy Inference System	Estimation
2016	[70]	Technical & Fundamental Indicator	DE: Differential Evolution Algorithm	Classification
2016	[71]	Technical Indicator	SVM: Support Vector Machine AB: Adaptive Boosting Algorithm GA: Genetic Algorithm	Classification
2017	[72]	Technical Indicator	LR: Logistic Regression GBM: Gradient Boosting Machine	Classification
2017	[56]	Technical Indicator	SVM: Support Vector Machine k-NN: k-Nearest Neighbor	Estimation
2017	[73]	Technical Indicator	DNN: Deep Neural Network ANN: Artificial Neural Network	Estimation
2017	[74]	Technical Indicator	k-NN: k-Nearest Neighbors	Estimation
2017	[75]	Technical Indicator	CSA: Cuckoo Search Algorithm GA: Genetic Algorithm	Estimation
2017	[76]	Technical Indicator	WNN: Wavelet Neural Network	Classification
2017	[77]	Technical Indicator	PLR: Piecewise Linear Representation SVM: Support Vector Machine	Estimation
2017	[78]	Technical Indicator	ANN: Artificial Neural Network SVM: Support Vector Machine	Estimation
2017	[55]	Technical Indicator & News Data	ANN: Artificial Neural Network DT: Decision Trees SVM: Support Vector Machine	Dataset Analysis
2017	[79]	Technical Indicator, News Data & Social Media	SVM: Support Vector Machine TeSIA: Tensor-based learning approach proposed CMT: Coupled Stock Correlation	Estimation
2017	[80]	Technical Indicator & Social Media	SVM: Support Vector Machine MLP: Multilayer Perceptron	Estimation
2018	[10]	Technical Indicator	SVR: Support Vector Machine Regression	Estimation
2018	[12]	Technical Indicator & News Data	RNN: Recurrent Neural Networks	Estimation
2018	[81]	Technical Indicator	RNN: Recurrent Neural Networks LSTM: Long Short-Term Memory	Estimation
2018	[82]	Technical Indicator	ANN: Artificial Neural Network	Estimation
2018	[83]	Technical Indicator	SVM: Support Vector Machines RF: Random Forest GB: Gradient Boosting AB: Adaptive Boosting Algorithm	Classification
2018	[41]	Technical Indicator	DT: Decision Tree SVM: Support Vector Machine MLP: Multilayer Perceptron Neural Network LSTMNN: Long Short-Term Memory Neural Network	Classification
2018	[84]	Technical Indicator & News Data	MKL: Multiple Kernel Learning	Classification
2018	[29]	Technical Indicator	SVM: Support Vector Machine	Classification

2018	[85]	Technical Indicator & News Data	SVM: Support Vector Machine LR: Logistic Regression	Classification
2018	[86]	Technical Indicator & News Data	DRSE: Deep Random Subspace Ensembles	Classification
2018	[13]	Technical Indicator & News Data	NNRE: Neural Network Regression Ensemble SVRE: Support Vector Regression Ensemble BRT: Boosted Regression Tree RFR: Random Forest Regression	Estimation
2018	[38]	Technical Indicator	RF: Random Forests	Dataset Analysis
2018	[87]	Technical Indicator	LSTM: Long Short-Term Memory CNN: Convolutional Neural Network	Estimation
2019	[3]	Technical & Fundamental Indicator	RF: Random Forest	Estimation
2019	[57]	Technical Indicator	NB: Naïve Bayes Classifier	Classification
2019	[88]	Technical Indicator	MR: Multiple Regression	Classification
2019	[89]	Technical Indicator	CNN: Convolutional Neural Network	Estimation
2019	[33]	Technical Indicator	CNN: Convolutional Neural Networks	Classification
2019	[30]	Technical Indicator	LSTM: Long Short-Term Memory	Association
2019	[59]	Technical Indicator	SVM: Support Vector Machine	Estimation
2019	[90]	Technical Indicator	CNN: Convolutional Neural Networks	Estimation
2019	[14]	Technical Indicator, News Data & Social Media	NB: Naïve Bayes (NB) SMO: sequential minimal optimization k-NN: k-nearest neighbour LWL: Locally Weighted Learning ASC: Attribute Selected Classifier PART: Partial C4.5 Decision Tree MLP: Multilayer Perceptron RF: Random Forest DT: Decision Table	Classification
2019	[40]	Technical Indicator	LR: Logistic Regression SVM: Support Vector Machine RF: Random Forest	Classification
2019	[91]	Technical Indicator	LR: Linear Regression SVM: Support Vector Machine CART: Classification and Regression Tree RF: Random Forest NB: Naïve Bayes Classifier XGB: eXtreme Gradient Boosting Algorithm MLP: Multilayer Perceptron DBN: Deep Belief Network SAE: Stacked Autoencoders RNN: Recurrent Neural Network LSTM: Long Short-Term Memory GRU: Gated Recurrent Unit	Dataset Analysis
2019	[28]	Technical Indicator	ANN: Artificial neural network SVM: Support Vector Machine RF: Random Forest LSTM: Long short-term memory	Estimation
2019	[92]	Technical Indicator & News Data	SVM: Support Vector Machine RF: Random Forest	Classification
2019	[58]	Technical Indicator	SVR: Support Vector Regression	Estimation
2019	[45]	Technical Indicator	LSTM: Long Short-Term Memory	Estimation
2019	[93]	Technical Indicator & News Data	SVR: Support Vector Regression	Estimation
2019	[94]	Fundamental Indicator	ELM: Extreme Learning Machine BPNN: Back Propagation Neural Networks SVR: Support Vector Regression	Estimation
2019	[8]	Technical Indicator & News Data	DNN: Deep Neural Networks	Estimation
2019	[7]	Technical Indicator	CNN: Convolutional Neural Network LSTM: Long Short Term Memory	Estimation
2019	[36]	Technical Indicator	XGB: Extreme Gradient Boosting	Classification
2019	[95]	Technical Indicator	SVM: Support Vector Machine	Estimation
2020	[96]	Technical Indicator	ANN: Elman Neural Network	Estimation
2020	[27]	Technical Indicator	SVM: Support Vector Machine PNN: Perceptron Neural Network	Estimation

			LR: Logistic Regression	
2020	[97]	Technical Indicator	MLP: Multilayer perceptron CNN: Convolutional neural networks LSTM: Long Short-Term Memory RF: Random Forest	Estimation
2020	[37]	Technical Indicator	kM: k-Means Clustering SVR: Support Vector Machine RF: Random Forest	Clustering
2020	[5]	Technical Indicator, News Data & Social Media	GNB: Gaussian Naïve Bayes MNB: Multinomial Naïve Bayes SVM: Support Vector Machine LR: Logistic Regression MLP: Multilayer Perceptron k-NN: k-Nearest Neighbor CART: Classification and Regression Tree LDA: Linear Discriminant Analysis AB: Adaptive Boosting Algorithm GBM: Gradient Boosting Classifier RF: Random Forest Classifier ET: Extra Tree	Classification
2020	[98]	Technical Indicator	k-NN: k-Nearest Neighbors	Classification
2020	[99]	Technical Indicator	RNN: Recurrent Neural Networks	Classification
2020	[15]	Technical Indicator & News Data	LSTM: Long Short Term Memory	Classification
2020	[17]	Technical & Fundamental Indicator	SVM: Support Vector Machine RF: Random Forest	Dataset Analysis
2020	[52]	Technical Indicator	DNN: Deep Neural Network SVM: Support Vector Machine RF: Random Forest LR: Logistic Regression	Classification
2020	[16]	Technical & Fundamental Indicator	LSTM: Long Short-Term Memory DNN: Deep Neural Network	Estimation
2020	[9]	Technical Indicator	LR: Logistic Regression MLP: Multilayer Perceptron RF: Random Forest XGB: Extreme Gradient Boosting LSTM: Long Short-Term Memory Network	Estimation
2020	[43]	Technical Indicator	MLP: Multilayer Perceptron LSTMNN: Long Short-Term Memory Neural Network CNN: Convolutional Neural Network	Estimation
2020	[100]	Technical Indicator	MLP: Multilayer Perceptron	Estimation
2020	[101]	Technical Indicator	DT: Decision Tree RF: Random Forest AB: Adaptive Boosting Algorithm XGB: eXtreme Gradient Boosting SVM: Support Vector Machine NB: Naïve Bayes k-NN: K-Nearest Neighbors LR: Logistic Regression ANN: Artificial Neural Network RNN: Recurrent Neural Network LSTM: Long short-term memory	Estimation
2020	[102]	Technical Indicator	HNN: Hybrid Neural Network	Estimation
2020	[103]	Technical Indicator	LSTM: Long short-term memory	Estimation
2020	[35]	Technical Indicator	ANN: Artificial Neural Network RF: Random Forest SVM: Support Vector Machine	Classification
2020	[104]	Technical Indicator	DT: Decision Trees SVM: Support Vector Machine ANN: Artificial Neural Network	Classification
2020	[31]	Technical Indicator	DQN: Deep Q-Network	Estimation
2020	[47]	Technical Indicator	RNN: Recurrent Neural Networks	Classification
2020	[25]	Technical Indicator	CNN: Convolutional Neural Network	Estimation
2020	[46]	Technical Indicator	RCN: Rough cognitive network	Classification