A LITERATURE REVIEW ON MACHINE LEARNING METHODS AND DATA FOR FORECASTING STOCK MARKETS

Abu Salim
Department of Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan, 45142, Kingdom of Saudi Arabia.
email: abu_salimak@yahoo.com

Aasif Aftab
Department of Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan,45142, Kingdom of Saudi Arabia,
E-mail- aaftab@jazanu.edu.sa

Khalid Ali Qidwai
Department of Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan, 45142, Kingdom of Saudi Arabia
Email: khalidqidwai@gmail.com

Mohammad Haseebuddin
Department of Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan, Kingdom of Saudi Arabia, Email: mhaseebuddin@jazanu.edu.sa

ABSTRACT:

In this work, we systematically reviewed the research on machine learning's potential for predicting the stock market and other financial markets. This literature review primarily focuses on the financial markets and the different kinds of input utilized by the machines learning algorithms used to predict the financial markets. For this, we combed over 138 papers published between 2000 and 2019. The primary contributions of this review are (1) a thorough analysis of the data used for the prognostications (including that of the markets and stock indices in the preconceptions and the 2173 exceptional relevant factors used for share price predictions, which can include candlestick patterns, macro - economic factors, and basic indicators) and (2) a thorough analysis of the machine learning techniques and variants used for the predictions. Included as well is a bibliometric study of these journal papers, which ranks articles according to how often they have been cited in other publications.

Keywords: Financial market, Data mining, Predictive performance, Classification, Regression, Stock market prediction.

DOI Number: 10.14704/nq.2022.20.13.NQ88214 **Neuro Quantology 2022; 20(13):1716-1734**

1. INTRODUCTION

The public's level of concern has skyrocketed during the previous several decades (Badolia, 2016). Investors act on the market with the hopes of generating a return over the duration of their investment horizon, therefore it's not surprising that everyday stock exchange transactions contain assets worth billions of dollars (Hoseinzade 2019). If a private or institutional investor, for instance, could accurately forecast the market's behavior, they may get a higher return on their money while taking on less risk. In light of this, it is reasonable to employ machine learning and other kinds of artificial intelligence to create accurate models for stock market forecasting. Some published studies (Sedighi et al., 2019) claim that the complicated forecasting models and algorithms they developed may accurately predict stock market movements, increasing the likelihood of financial gain for investors (Armano et al., 2005). Forecasting the stock market is a vital yet challenging task, and this is well recognized in the world of finance.

One cannot predict future price market movements by evaluating past ones, according to the efficient market hypothesis (EMH) presented by Fama (1970). Example: the EMH classifies market efficiency as either weak, moderate, or strong (Atsalakis & Valavanis, 2009a). The current stock price does not provide any new information to a time series beyond that which is already there in the prices, according to the idea of weak-form market efficiency (Fama, 1970). For this reason, the bare-bones form of EMH states that technical analysis cannot generate a higher return than a buy-and-hold strategy (Fama, 1965). Second, according to the semi strong EMH, stock prices fully reflect all available public information.

Semi-strong market efficiency holds that an investor cannot consistently outperform the market using publicly accessible information like basic data. It follows that active management that incorporates all available data may not outperform passive management in terms of risk-adjusted returns. In opposed to the semi strong version of the EMH, the strong form of the EMH asserts that insider information is already reflected in stock prices. This means that no investor, not even those who possess access to confidential information, can consistently achieve higher returns than the market (Fama, 1965). Therefore, according to the extreme version of the EMH, it is impossible to anticipate future stock market returns (Timmermann 2004). To illustrate how severe the strong form of the EMH may be, even Eugene recognized that one would not expect that insider knowledge (such as that of business leaders) cannot be leveraged to achieve higher expected profits.

As time goes on, more and more people begin to question the efficient market theory, which claims that asset values should accurately represent fundamentals (Daniel et al., 1998). Market anomalies that contradict the efficient market hypothesis include overreaction and underreaction in the financial markets, progress in the near term followed by a reversal in the long run, and getting out of control in asset prices (Malkiel & Mullainathan, 2005). Both overreactions and underreactions are common, according to the available studies. Institutional investors may be able to offset the irregularities brought about by less seasoned investors (Shiller, 2003). Yet, the inconsistencies can defy explanation by a model predicated on the underlying rationality of investors. Because of this, Bond and Thaler (1990) developed behavioral finance, which questions the idea of perfectly rational investors by factoring in human psychological biases including loss aversion, overreaction, and overconfidence (Lo, 2004). For a detailed account of how the efficient market hypothesis came to be, read Lim and Brooks (2011). Although market participants should be aware of the possibility of market anomalies, it is not surprising that many of them base their predictions of future stock prices on knowledge of past market prices, information specific to a company (such as past earnings and profits), and other variables. In addition, investors often think that recent gains will continue, since past gains can be a barometer of investor sentiment. It seems reasonable to try to predict the stock market using historical data given these assumptions and the possibility of market anomalies.

Researches that aim to predict the stock market typically employ both fundamental analysis and technical analysis (Lam, 2004). The same information can be used to predict the performance of a stock index, which is a group of stocks from different companies, as well as information on the competitive landscape, such as economic growth, trade, exchange rates, or interest rates, all of which are thought to have an impact on the businesses included in the index. The technical analysis method, on the other hand, predicts future stock price movements based on the examination of historical data, including both stock prices and trading volumes (Lohrmann & Luukka, 2019). Academics are drawn to these methods because they provide fresh and useful approaches to the age-old problem of forecasting the behavior of the financial markets. Using historical data is a cornerstone of the stock market forecasting method. Each of the aforementioned methods of analysis relies on carefully placed data (variables) within a certain time range.

Here, we conduct a thorough literature review of financial paper aims to design, focusing on the methodologies, data sources, and statistical and machine learning methods that have been used in the field. Therefore, the purpose of this work is to advance knowledge in stock market forecasting by conducting a systematic literature review. Our research questions were formulated beginning with this point of view. First, what information (variables, variable kinds, and time periods) is used? secondly, what sorts of artificial intelligence (AI) and machine learning (ML) approaches were taken to analyze data and make stock market predictions? After considering the recommendations offered by Snyder (2019), we conducted a thorough research study for this investigation following the methods described by Ahmad et al. (2018). (2013).

As a whole, 138 articles were located that ranged in publication date from 2000 to 2019. Furthermore, we utilized a number of standards for acceptance and rejection. To further organize the text, we did as follow: Section 2 provides a high-level overview of the data (and their attributes) and machine learning strategies used in stock market forecasting. Some related studies are discussed in Section 3. Section 4 provides a brief overview of the research methods used in this literature review. The results of the review study are presented and discussed in Section 5. Section 6 provides final thoughts on the article's findings and evaluations.

2. GENERAL OVERVIEW OF THE STOCK MARKET PREDICTION DATA AND MACHINE LEARNING TECHNIQUES

2.1. Data sources

A number of elements, such as technical analysis, financial variables, and economic determinants, have been recognized as important in the literature for understanding stock price changes. Studies that employed a wide range of input data to develop their stock market forecasts were done since there is no agreement on which factors are most important. As can be seen in Fig. 1, we categorized all the input variables found in the reviewed literature into four overarching groups with several subdivisions. Technical indicators are often employed as input variables in prediction studies due to their significance in offering buy and sell recommendations for equities. We classified all technical indicators into two groups: "basic technical indicators" and "extra technical indicators." In Section 5.3, we discuss the stock market indexes and factors that have been the subject of previous research. Successful stock market forecasting studies have used fundamental technical indicators as input data (Chun & Park, 2005); other studies have adopted different technical indicators (Chang & Wu, 2015).

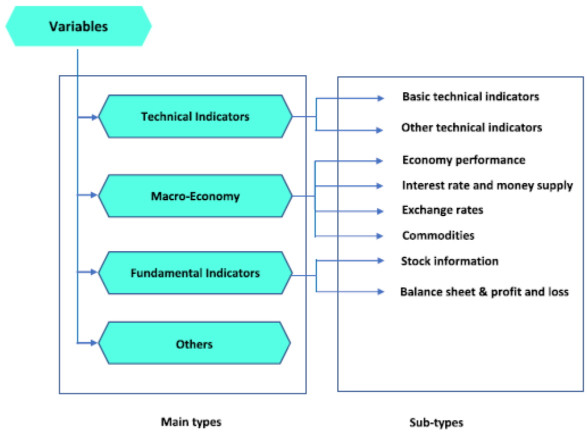

Fig. 1. Variable categories for stock price and return predictions [1]

Here, we divided things up into four categories: commodities, economic performance, interest rates, and money supply. To further refine stock market predictions, fundamental indicator-related variables have also garnered attention (Barak et al., 2017). Such factors are divided into two categories for the sake of discussion.

2.2. Machine learning

Machine learning is predicated on the concept that new knowledge may be gained from previously collected data (Kubat, 2017, p. 1). Supervised learning is the most common method of machine learning training used for stock market prediction applications. In Fig. 2, we see a comprehensive depiction of the steps involved in employing supervised learning to make predictions about the stock market. Time series data (such as stock price and/or return) and/or pertinent information is selected as the first stage (such as financial news). In a classification problem, the target class is something you are already familiar with or can easily anticipate. Pre-processing is required for the supplementary data, which includes cleaning it of any irrelevant or extraneous details (such as identifiers). Then, technical indicators can be derived from the core time analysis, such as the close price information. The cleansed data, that may include technical indicators, undergoes scaling and dimension reductions (namely feature selection, edge detection, and feature construction) to get significant variables and filter out unnecessary ones. Pre-processed data is commonly used to make accurate predictions (Chen et al., 2019). While this step is essential, it is portrayed as optional in Fig. 2 since it is not only creator but also subject-dependent. After gathering raw data, the following step is to select a well-established method for machine learning or create a whole new one in order to forecast the target variable.

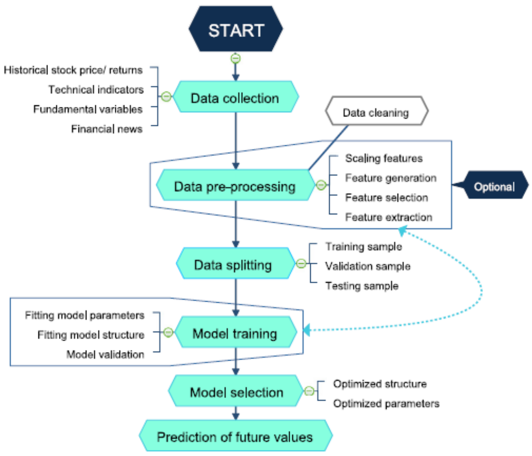

Fig. 2. Stock market forecasting using supervised learning workflow. [1]

As a means of accomplishing this, input data are often divided into training data (used for really building the classifier with a defined set of parameters and structure), validation data (used for actually assessing the performance of all trained models and identifying the best model structure and parameters), and test data. The most basic kind of feature selection (i.e., filter approaches) is included in data preparation since it may be used irrespective of the learning process. It might be linked to model training, though, either by including the learning computation performance into feature selection (the integrated method) or by using the learning algorithm's performance to carry out feature selection (the wrapper technique) (Guyon & Elisseeff, 2003). To illustrate the potential connection between feature selection and model training, a dashed line connects the training stage to the pre-processing step. The next stage is prediction, which may be carried out using either the trained regression model or the taught classification model.

Recently achieved success in applying deep learning techniques to stock market forecasting has re-elevated these techniques to the forefront. Without depending on life experiences or economic assumptions, deep learning is a type of deep learning that could extract useful characteristics from challenging and noisy data. Stock market price and return forecasting has seen widespread application of dnns (Singh & Srivastava, 2017), including convolutional layers and long bad memory networks (LSTM) (Fischer & Krauss, 2018).

3. RELATED WORK

There has been a rise in the number of review studies published across several disciplines, including business and finance (Henrique et al., 2019), computer science (Ahmad et al., 2018), social science, medicine (Klemm et al., 2003), and engineering (Ambreen et al., 2018). Financial market forecasting models have received scant attention in the literature. More than a hundred publications were surveyed for their relevance to the topic of stock market estimation techniques in Atsalakis and Valavanis's 2009 review. (2009b). This article examines the application of neural networks and neuro-fuzzy systems. Rather et al. (2017) conducted an analogous literature analysis, this time focusing on the ways in which mathematical models and AI were used to make stock market predictions and investment selections. Farias Nazário et al. conducted a literature survey on the topic of stock market forecasting with technical analysis indicators (2017). They did a literature study on technical analysis and provided recommendations for new studies. An up-to-date survey of machine learning techniques for stock market forecasting is available in (Henrique et al., 2019).

A bibliographic study of 57 selected pieces of research is presented, discussing the most cited works, authors, and co-citation rates. In their extensive study and review of the research on stock market prediction, Gandhmal and Kumar (2019) drew from more than fifty research. They effectively classified the studies by their prediction methods, year of publication, performance standards, and analysis software. In addition, Shah et al. provided a summary analysis and classification of models used to forecast the stock market (2019). Some recent evaluations have tried to cover more of the literature on the subject. Bustos and Pomares-(2020) Quimbaya's comprehensive review of stock market prediction techniques include 52 publications during 2014-2018. This article provides a summary of recent developments in machine learning techniques, including deep learning, text mining, and ensemble approaches. During the same time period, Jiang (2021) examined deep learning models that have been employed to make predictions about the stock market. Some potential directions for further study were also discussed, along with a brief overview of the data and data processing methods used. By looking at 30 articles from scholarly journals and conferences, Kumar et al. (2021) conduct a comprehensive landscape analysis of the machine learning algorithms, quality measures, datasets, and publications utilized in stock market prediction research.

Fig. 3. The literature review process [1]

4. RESEARCH METHODOLOGY

4.1. Planning

Doing a systematic literature examination requires specifying the review procedure in order to gather primary studies and reduce bias (such as biasness) in our study (i.e., the entire strategy). Since Kitchenham's method has been demonstrated to be beneficial in earlier evaluations, we applied it (2004). Includes methods for conducting the preliminary research and conducting the actual review. Together, we came up with a plan for our systematic literature search research, which we carried out according to the procedure depicted in Fig. 3.

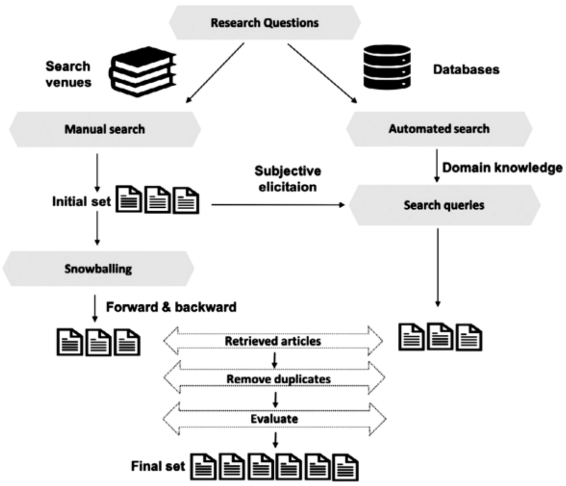

Fig. 4. The mechanism of the search operation.[1]

4.2. Search strategy

Finding the best publications that directly address these issues was the primary objective of the search technique. This review study's literature searches were conducted in two distinct periods. We started by doing a "manual search," or selecting the first set of items to read from a set of results. We utilized this first set of articles to kick off a snowball effect, following Wohlin's methodology (2014). Second, we used a technique very similar to that proposed by Kitchenham and Brereton and conducted a "automated search" for other relevant publications (2013). This strategy involved identifying the specific scientific databases to query and then developing a set of overarching search terms to be used in those queries. Therefore, Fig. The search process is shown in detail in 4.

4.3. Search queries

The importance of our study is emphasized by the search queries we have specified. To come up with a working definition, we drew from our prior knowledge and experience in the subject, as well as the titles, keywords, and abstracts we found throughout our manual search. We opted to analyze the two terms separately since we discovered that the relevant literature used either "prediction" or "forecasting" in their definitions. It is common knowledge that a query including the phrases "Stock price" and "Stock return" would provide predictive or forecasting results for just one of the terms. The phrase "Machine Learning" was selected as the final term of the specified inquiry since the relied-upon articles are expected to be based in forecasting models that employ the use of machine learning techniques. However, it was found that when "AI" was used alone, without "Machine Learning," a more comprehensive set of publications on the topic was generated. Queries executed against various data sources to allow automated searching.

4.4. Inclusion and exclusion criteria, and study selection

Following rigorous inclusion and exclusion criteria, we selected the articles most relevant to our study. Articles that were presented at conferences, were never published, or were included in a report, thesis, or book were not included. One of the reasons we read the story was because we could get the whole thing online. Articles published in language other than English were also discarded.

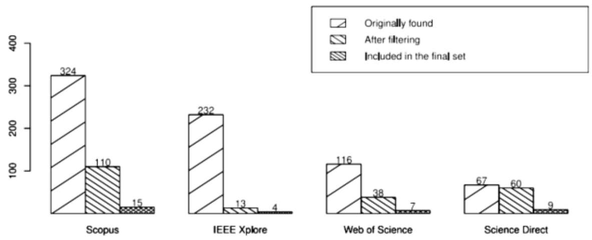

Fig. 5. Databases and the corresponding number of articles in each stage. [1]

4.5. Data extraction

Our decision to summarize just certain pieces of data/qualitative characteristics was guided by this literature review's aims. Relevant information was found and extracted based on the predefined attributes after reviewing each item in its full. For additional analysis, the information was placed into a Microsoft Excel spreadsheet. Note that several attributes were further separated (using extra columns) to accommodate all required data. When everything was added together, we came up with a total of 79 separate characteristics.

Variables come in many forms, the most common of which are addendum, macro - economic variables, fundamental measurement items, and others. As a result, we gathered information on data elements referenced in the literature as being useful for predicting stock market performance. To learn this, we assigned each study a number of "1" if a certain feature (predefined, as specific as feasible, such as "basic moving average (5 periods)") was used, and a value of "0" otherwise. When there was insufficient data to reliably differentiate between several possible configurations of a characteristic used in a study, the number of configurations was utilized instead.

4.6. Data synthesis

The purpose of this literature review is to analyse and summarize the data presented in the chosen articles in order to provide answers to the study's research questions. To do this, we employed both a descriptive analysis of the statistical data and a thematic synthesis of a primary analysis. The information within was subjected to both descriptive and analytical scrutiny.

5. RESULTS OF THE LITERATURE REVIEW

Results from our comparative analysis are presented below, considering the aforementioned research foci. Extraction of data from this primary research (138 in total) was used to conduct the analysis. Three reviews' findings are presented and discussed in terms of (1) bibliographical information, (2) data and their quality, and (3) machine learning algorithms that were developed or deployed.

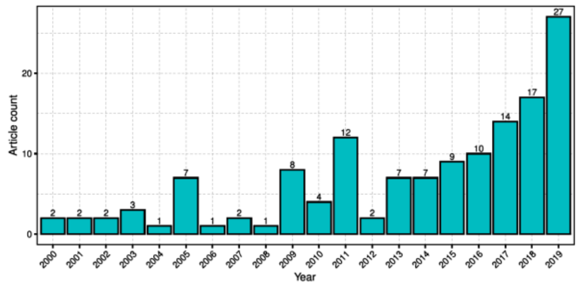

Fig. 6. Number of included articles by year [1]

5.1. A review of bibliographic information

For each of the issues brought up in the pilot research, we conducted a bibliographic analysis and present a case study below. Thereafter, the selected articles are broken down by year, journal, most-used words, and citation count. First, we give a brief overview of the research that inspired our inquiry.

5.1.1. Overview of extracted data

To compile this overview, we combed through 52 academic journals spanning 20 years and extracted 138 primary papers.

5.1.2. Publications by yea

Figure 6 depicts the total series of papers produced over the years. Most of the studies found here were recent articles, making this topic set extremely relevant. There will have been more than half as many publications in 2015 and 2019 than there were in the prior 15 years combined. The number of studies using machine learning to predict the stock market looks to be growing at an exponential rate.

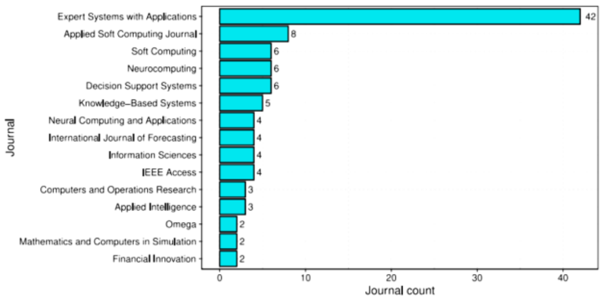

Fig. 7. Article counts by most relevant journals [1]

5.1.3. Publications by journal

Figure 7 displays the most pertinent journals to the research. All told, 52 journals had items selected for inclusion. We only show the 16 magazines (out of 138 total) that have multiple articles from our dataset. According to this metric, Expert Systems with Implications is the gold standard for stock market forecasting articles. Nearly 31% of the reviewed articles fall into this category. Journals such as Applied Soft Technology, Soft Data processing, Direct learning model, Management Information Systems, and Assumptions About the world account for around 22% of the total. This study breaks down which periodicals have included the biggest articles on using machine learning to predict the stock market.

5.1.4. Variation of keywords

The inclusion of a list of keywords in a scientific paper is crucial since it makes the content searchable and facilitates its easy accessibility to other academics. Here we can observe the 20 most often used author keywords in Fig. 8. The fact that some of the most popular search phrases are available in both their singular and plural forms (such as "CNN architecture" and "neural networks") implies that using both forms might improve the efficiency with which you find relevant items.

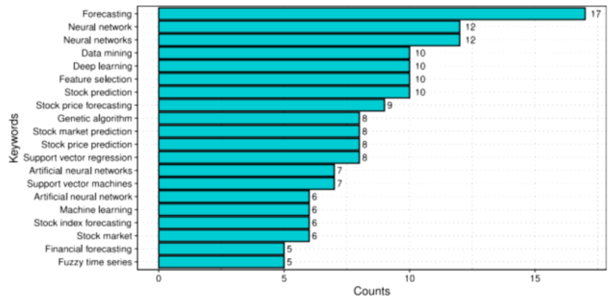

Fig. 8. Top authors and keywords used most frequently to count articles. [1]

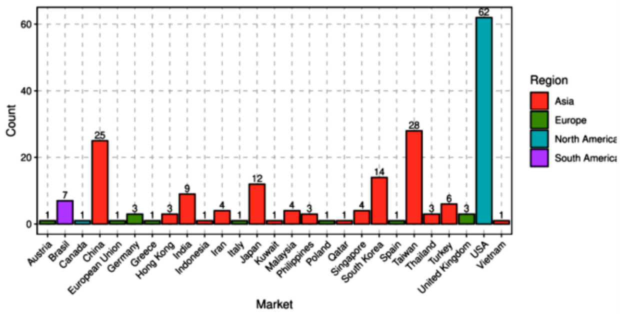

Fig. 9. Frequency by market. [1]

5.2. Markets, indices, and stocks

This paper's forecasting attempts centre on distribution of wealth and income on securities exchanges and standards that measure the quality of a selection of individual stocks. Examples of stock indices include the Standard & Poor's 500, which tracks the value of 500 of America's largest publicly traded companies (including "Apple Inc."). Research articles representing 26 separate markets and a single EU-wide index were used in this analysis. The number of pieces that discussed markets or specific stocks is seen in Figure 9. Only four of the world's seven main regions are represented in our data: Asia, Europe, North America, and South America. Sorry to apologize, Antarctica goes without saying, but neither Africa nor Australia are mentioned. Only one book makes specific mention of Canada, suggesting that the entirety of North America is considered to be inside the United States. Furthermore, South America is represented entirely by Brazil. However, Asia's 16 distinct markets make it the largest cluster in our analysis. Europe is represented by its European Union and its seven individual markets. North America, with a total of 63 insureds in the literature, clearly ranks second among the areas, i.e. behind Asia, which has 119 occurrences. This is despite the fact that the United States is the most widely covered market overall. Only twelve references to Europe and seven to South America can be found in the literature on stock market forecasting. Figure 10 shows the frequency with which various exchanges and indices were mentioned, as well as the marketplaces in which they were employed, throughout our research. There are 45 distinct exchanges and indexes in the studied literature (see Appendix for a comprehensive list with full names and markets).

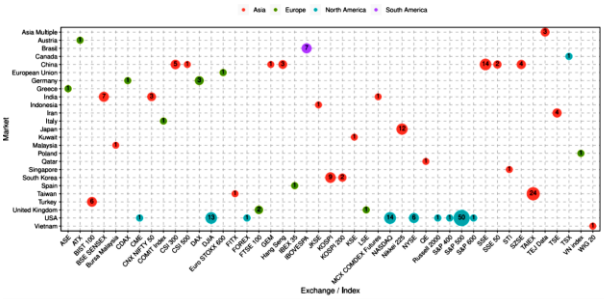

Fig. 10. Frequency of exchanges/indices (by market). [1]

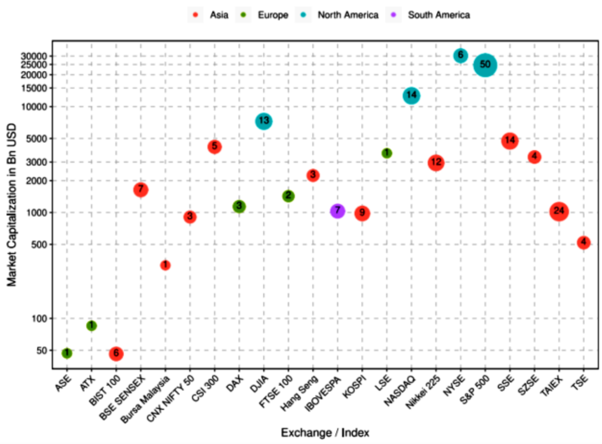

Fig. 11. Frequency of indices by market capitalization. [1]

A total of 22 indices are included in this comparison, the most prominent of which are the four U.S. indices assessed in the study publications. The value of the Dow Jones Industrial Index is $7.252 quadrillion dollars, while the value of the NYSE Composite Index is $30.21 billion dollars. None of the European or Asian markets that were analyzed are above the range's ceiling. The size of the market is compared to how often markets are discussed in books and other media. The Pearson correlation is 0.545, and the Spearman rank correlation is 0.503, indicating a moderate positive association between the number of occasions these indices/exchanges are referenced in academic publications and their total market capitalization.

Since this is the case, it would appear that academic studies pay more attention to the major stock markets than they do to the smaller ones. However, this isn't true of all markets. Istanbul Stock Exchange (BIST 100) is mentioned just six times despite being one of the smaller exchanges with yearly revenue of $46.2 billion. However, despite being one of the largest stock exchanges considered here with a market capitalization comparable to 3.625% of USD, the London Stock Exchange's principal stocks market earns just one mention. However, it is also evident that the largest stock market is not always the one that has been studied the most. The S&P 500 [USA] is cited eight times as often as the NYSE Composite Index [USA], even though the former is the more generally watched indicator of the North American stock market. 4 Although the TAIEX (Taiwan) is quoted nearly twice as often as the SSE Composite Index (China), the latter is the most often cited Asian index.

You'll find a comprehensive breakdown of all the stocks that had a role in our research here, whether as input to a stock market prediction model or as the final target of that model's predictions. The North American stock market accounts for a disproportionately large share of the global total, followed by the Asian stock market. European and South American stocks were also the subject of research in the scholarly literature. By these measures, it is evident that North American stocks (50.8%, excluding "Ceased to Exist") are more widely discussed in the academic literature than their Asian counterparts (42.2%).

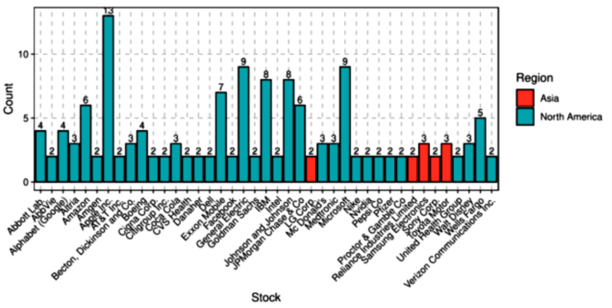

Fig. 12. Frequency of the Top 41 stocks. [1]

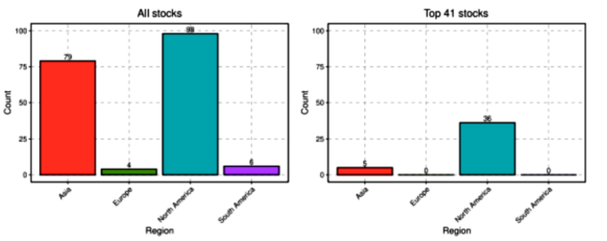

Fig. 13. All stocks vs the Top 41 stocks by region. [1]

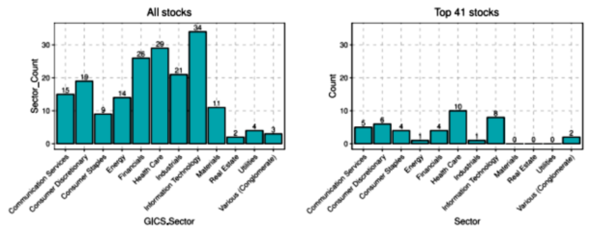

Fig. 14. All stocks vs the Top 41 stocks by GICS sector. [1]

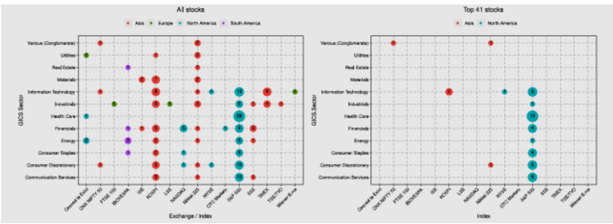

Fig. 15. All stocks vs the Top 41 stocks by index/exchange, GICS sector and region [1]

5.3. Variables for stock and index prediction

In all, this study incorporates the findings from 2173 individual parts and 138 different scholarly papers. Variables were sorted using the categories Technical Indicator, Macro-Economy, Core Indicator, and Other. In the context of economic variables of time series data, such as stock prices, a technical indicator is a pecuniary component. The relative performance index (RSI), moving averages, probabilistic oscillators, etc., are all examples of additional, on average greater complex technical indicators beyond the closing, open, high, and low prices and volume. The "Macro-Economy" group of variables has several potentially fascinating clusters. Data on economic performance includes things like gross domestic product (GDP) and rate of interest and money supply things like treasury bill rates and other variables tracking the flow of cash; exchange rate things like the US dollar and the Euro; and commodity things like gold and crude oil. Variables linked to financial models (such beta in the capital asset pricing model (CAPM) or excess return) coexist with those related to machine learning since the "Other" category does not further identify the variables it contains. Bag of words, linear regression line, etc. Figure 16 shows the frequency with which each variable appears in each group.

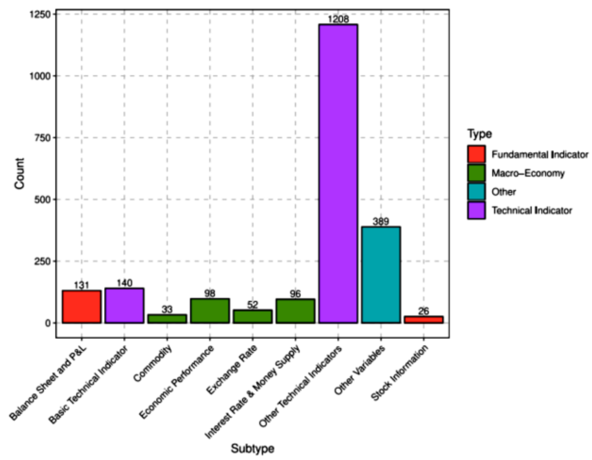

Fig. 16. Number of variables by sub-type [1]

5.3.1. Technical indicator variables

The basic technical indicators contain 140 different variables that are direct reflections of crucial price and volume data from stock markets. A few examples of such variables include the open price, the tonnage traded at that price, the maximum and minimum prices (the highest and lowest prices at which a stock was moved within a certain time period), and the close price (the last recorded price at which a stock was purchased on a given trading day). It's clear that there are a few primary factors that are used far more often than the others. These contain the open and close prices, high and low prices, and volume for a certain time period, with a latency of 1. For accurate forecasting, it's best to use variables like the amount of trades that have lower lag values, shorter time durations, and more easily available information.

5.3.2. Macro-economic variables

Exchange rates

Across the papers we looked at, we found 53 unique factors associated with different pairs of currencies. Including the US dollar, these exchange rates feature a total of 12 distinct currencies: 11 single currencies and 1 basket of currencies. Dollar Index measures the greenback vs a group of six other currencies (Euro, Japanese Yen, British Pound Sterling, Swiss Franc, Canadian Dollar, and Swedish Krona).

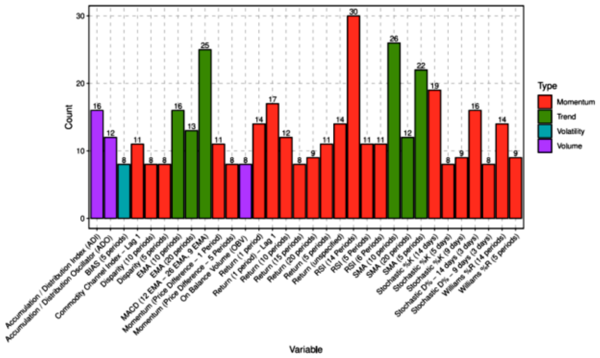

Fig. 17. Top 30 technical indicators. [1]

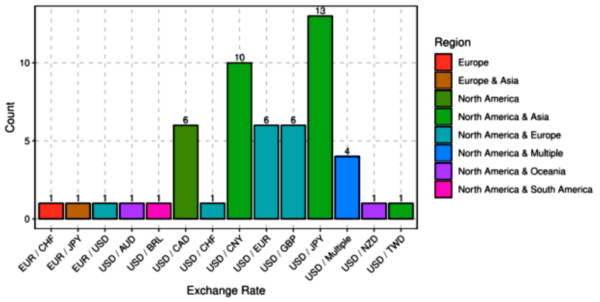

Fig. 18. Exchange rates. [1]

Commodities

Metals of high value and energy sources like natural gas and oil are part of this category. The United States Industrial Index, the most widely used index or ETF, reflects 56.6% of them (ETFs). Commodity futures trading accounts for 36.4% of all trading, only behind the more common practice of utilizing commodity prices (5.9%). After doing this investigation, we can say with confidence that oil production is the most often referenced commodity, coming 20 times (42.6% of the time). In addition, 14 references (29.8%) can be found in the literature on stock market forecasting that involve gold. We looked at the literature on stock market forecasting and found that three criteria were frequently mentioned. Percentage movements in crude oil, metal, and gold with a 1-period lag are shown below.

Economic performance

This group of macroeconomic indicators includes 98 separate variables that may be broken down into roughly five categories. The following five main economic indicators are taken into account: (1) "National Productivity," Second, the "Trade"; third, "Revenue & Invest," which includes private and government revenue as well as capital inflows and government expenditures; fourth, the "Labor Market;" and fifth, the "Sector" performance of selected economic sectors. The "Labor Market" subcategory is the youngest of the group, with only two variables (2.0%). At least twice among the articles we analyzed did this set of factors appear. It has been repeatedly proved that just a small subset of variables is really used in a large number of articles across all journals.

5.3.3. Other variables

There are two more sets of varying factors to consider in addition to those already presented. The first one is a data warehouse for a stock market prediction model, where stocks and market parameters are stored. Companies such as (Microsoft, ExxonMobil, General Electric, J&J, P&G) are examples. To predict the performance of the 10 largest stocks in the KOSPI 200 and the S&P 500, respectively, Baek and Kim (2018) employ a methodologically similar strategy. However, other studies have only used market indices, not index components or other large-cap shares, as their data source. Additionally, standard technical markers and returns from six major equity markets were also included.

5.4. Machine learning: a review

The major reason we conducted this literature study was to provide a comprehensive evaluation of the existing machine learning-based stock market forecast algorithms. Each prediction model in our database of reviewed articles was classified as either a regressed prediction, a classification forecasting, or a clustering prediction. The majority of the studies included in the graph are regression analyses. Despite their progress, clustering models still only account for around 1.4% of all scholarship, whereas studies focused on categorization account for over 44%. (Only two researches).

6. STUDY LIMITATIONS

We looked at the two main types of machine learning, supervised and unsupervised. While statistical forecasting methods such as ARIMA and GARCH were briefly discussed in a few studies (maybe as alternatives to the proposed technique), they were not the focus of our investigation. MCDM approaches, which are predominant on profit planning but not in the realm of addiction, were not incorporated into our research. Finally, our focus on machine learning meant that we couldn't investigate a wide range of other soft-computing techniques, such as those employed in regression analysis. It is possible that pertinent research presented at conferences or written in peer-reviewed symposia or technical reports was overlooked because our search was limited to papers published in academic journals. It's possible that we missed some relevant articles because of the terms we used in our searches.

7. DISCUSSION AND CONCLUSIONS

In this research, we do a comprehensive literature assessment of stock market prediction articles that make use of machine learning. We looked at 138 publications from 2000-2019 published in academic journals. Highly cited papers, keyword distribution by year, and other insights were uncovered by analysing the collected data. The next step was an in-depth investigation of the financial data sets and their accompanying independent and dependent variables described in the literature review. The data analysis of markets, indices, and stocks showed that U.S. indices and stocks, especially the S&P 500 index, were the most studied. However, Asia was the most researched continent overall. In particular, the combination of Taiwan, China, and South Korea was weighed more heavily than the United States. Consequently, it should come as no surprise that the TAIEX (Taiwan) was the second most researched index in our analysis. We also discovered a modest link between the size of a stock index's market capitalization and the number of times that index was discussed in the literature, suggesting that larger indices are more often covered. When looking at specific equities, articles focused mainly on those related to the healthcare industry, the IT sector, and the discretionary spending sector of the economy. In addition, we analyzed 2173 distinct variables found in the primary sources.

Technical Indicator has the most variables (1348) of any category. We discovered that technical indicators accounted for the biggest percentage of all variables, on average about 80%, notably for lower dimensional data with less than 20 variables. However, there was a trend for "Macro-Economy" and "Fundamental Indicator" particular variables to make up a bigger proportion of the total set of variables as the collection grew larger. In our analysis, we found that studies with more than 500 variables tended to employ the bag of words method of text mining, in which pertinent keywords are selected from data such as financial news or tweets. Based on our research, we found that the ANN, the SVM, and fuzzy theory were the most popular machine learning-based prediction models for stock market forecasting. This result differs slightly from the data offered by Henrique et al. (2019), as they had not published on fuzzy theory-based approaches prior to this study. Moreover, we found that deep learning approaches have been widely discussed in the previous three years, as evidenced by the presence of 15 papers on the topic in the last year alone. In addition, the recent decade has seen a marked increase on the use of methods grounded in fuzzy set theory. When trying to enhance the parameter optimization of machine learning algorithms, GAs were also widely employed. In the field of feature extraction, principal component analysis (PCA) and the wavelet transform were the two most often used techniques. Based on a meta-analysis of previous studies, we know that deep learning's meteoric surge in popularity will continue well into 2021. The new study finds that all deep learning-based articles have used enhanced LSTM models to predict stock market variables. This research demonstrates that LSTM networks are superior to other types of deep learning (such as CNNs and RNNs), yielding more accurate predictions.

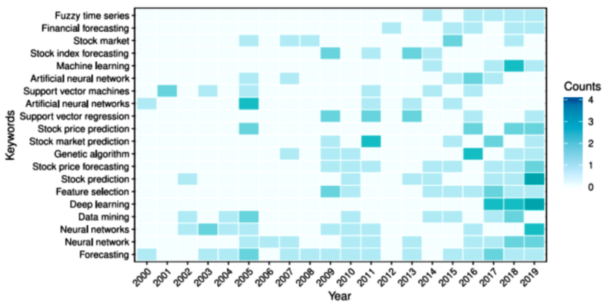

Fig. 19. Author's keywords dynamic by year [1]

REFERENCES

[1]. Mahinda Mailagaha Kumbure, Christoph Lohrmann, Pasi Luukka, Jari Porras. (2022). Machine learning techniques and data for stock market forecasting: A literature review, Expert Systems with Applications, Volume 197, 116659, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.116659.

[2]. Ahmad, M. O., Dennehy, D., Conboy, K., & Oivo, M. (2018). Kanban in software engineering: A systematic mapping study. Journal Of Systems And Software, 137, 96–113. http://dx.doi.org/10.1016/j.jss.2017.11.045.

[3]. Ahmadi, E., Jasemi, M., Monplaisir, L., Nabavi, M. A., Mahmoodi, A., & Amini Jam, P. (2018). New efficient hybrid candlestick technical analysis model for stock market timing on the basis of the support vector machine and heuristic algorithms of imperialist competition and genetic. Expert Systems With Applications, 94, 21–31.

[4]. Ahmed, Y. A., Ahmad, M. N., Ahmad, N., & Zakaria, N. H. (2019). Social media for knowledge-sharing: A systematic literature review. Telematics And Informatics, 37, 72–112. http://dx.doi.org/10.1016/j.tele.2018.01.015.

[5]. Altay, E., & Satman, M. H. (2005). Stock market forecasting: Artificial neural network and linear regression comparison in an emerging market. Journal Of Financial Management And Analysis, 18(2), 8–33.

[6]. Ambreen, T., Ikram, M., & Niazi, M. (2018). Empirical research in requirements engineering: trends and opportunities. Requirements Engineering, 23, 63–95.

[7]. Anbalagan, T., & Maheswari, S. U. (2015). Classification and prediction of stock market index based on fuzzy metagraph. Procedia Computer Science, 47(C), 214–221.

[8]. Anish, C. M., & Majhi, B. (2016). Hybrid nonlinear adaptive scheme for stock market prediction using feedback FLANN and factor analysis. Journal Of The Korean Statistical Society, 45(1), 64–76. http://dx.doi.org/10.1016/j.jkss.2015.07.002.

[9]. Araújo, R. D. A. (2011). Translation invariant morphological time-lag added evolutionary forecasting method for stock market prediction. Expert Systems With Applications, 38(3), 2835–2848.

[10]. Araújo, R. D. A., Oliveira, A. L., & Meira, S. (2015). A hybrid model for high-frequency stock market forecasting. Expert Systems With Applications, 42(8), 4081–4096.

[11]. Armano, G., Marchesi, M., & Murru, A. (2005). A hybrid genetic-neural architecture for stock indexes forecasting. Information Sciences, 170(1), 3–33.

[12]. Atsalakis, G. S., & Valavanis, K. P. (2009). Forecasting stock market short-term trends using a neuro-fuzzy based methodology. Expert Systems With Applications, 36(7), 10696–10707. http://dx.doi.org/10.1016/j.eswa.2009.02.043.

[13]. Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques - part II: Soft computing methods. Expert Systems With Applications, 36, 5932–5941. http://dx.doi.org/10.1016/j.eswa.2008.07.006.

[14]. Badolia, L. (2016). How can i get started investing in the stock market. Educreation Publishing, 2016.

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

[16]. Barak, S., Arjmand, A., & Ortobelli, S. (2017). Fusion of multiple diverse predictors in stock market. Information Fusion, 36, 90–

102. http://dx.doi.org/10.1016/j.inffus.2016.11.006.

[17]. Barak, S., & Modarres, M. (2015). Developing an approach to evaluate stocks by forecasting effective features with data mining methods. Expert Systems With Applications, 42(3), 1325–1339. http://dx.doi.org/10.1016/j.eswa.2014.09.026.

[18]. Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. North American Journal Of Economics And Finance, 47, 552–567. http://dx.doi.org/10.1016/j.najef.2018.06.013.

[19]. Bataa, E., Vivian, A., & Wohar, M. (2019). Changes in the relationship between shortterm interest rate, inflation and growth: evidence from the UK, 1820-2014. Bulletin Of Economic Research, 71, 616–640.

[20]. Bisoi, R., Dash, P. K., & Parida, A. K. (2019). Hybrid variational mode decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis. Applied Soft Computing, 74, 652–678. http: //dx.doi.org/10.1016/j.asoc.2018.11.008. B

[21]. odie, Z., Kane, A., & Marcus, A. J. (2009). Investments (8th Ed.). Irwin: McGraw-Hill.

[22]. Bondt, W. F. M. D., & Thaler, R. (1985). Does the stock market overreact? The Journal Of Finance, 40, 793–805.

[23]. Bondt, W. F. M. D., & Thaler, R. H. (1990). Do security analysts overreact? The American Economic Review, 80, 52–57.

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

[25]. Braun, V. (2006). Using thematic analysis in psychology. Qualitative Research In Psychology, 3, 77–101.

[26]. Bustos, S. M., Anderson, J. V., Miniconi, M., Nowak, M., Roszczynska-Kurasinska, M., & Brée, D. (2011). Pricing stocks with yardsticks and sentiments. Algorithmic Finance, 1, 183–190.

[27]. Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A systematic review. Expert Systems With Applications, 156, Article 113464.

[28]. Cagcag Yolcu, O., & Alpaslan, F. (2018). Prediction of TAIEX based on hybrid fuzzy time series model with single optimization process. Applied Soft Computing, 66, 18–33. http://dx.doi.org/10.1016/j.asoc.2018.02.007.

[29]. Cao, Q., Leggio, K. B., & Schniederjans, M. J. (2005). A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market. Computers And Operations Research, 32(10).

[30]. Cao, H., Lin, T., Li, Y., & Zhang, H. (2019). Stock price pattern prediction based on complex network and machine learning. Complexity, 2019.

[31]. Cao, L., & Tay, F. E. (2001). Financial forecasting using support vector machines. Neural Computing And Applications, 10, 184–192.

[32]. Cao, J., & Wang, J. (2019). Stock price forecasting model based on modified convolution neural network and financial time series analysis. International Journal Of Communication Systems, 32, 1–13.

[33]. Carosia, A. E. O., Coelho, G. P., & Silva, A. E. A. (2019). Analyzing the Brazilian financial market through portuguese sentiment analysis in social media. Applied Artificial Intelligence, 34, 1–19.

[34]. Carta, S., Consoli, S., Piras, L., Podda, A. S., & Recupero, D. R. (2021a). Event detection in finance using hierarchical clustering algorithms on news and tweets. PeerJ Computer Science, 7.

[35]. Carta, S. M., Consoli, S., Piras, L., Podda, A. S., & Recupero, D. R. (2021b). Explainable machine learning exploiting news and domain-specific lexicon for stock market forecasting. IEEE Access, 9, 30193–30205. http://dx.doi.org/10.1109/ACCESS.2021.3059960.

[36]. Chai, J., Du, J., Lai, K. K., & Lee, Y. P. (2015). A hybrid least square support vector machine model with parameters optimization for stock forecasting. Mathematical Problems In Engineering, 2015.

[37]. Chang, P. C., & Fan, C. Y. (2008). A hybrid system integrating a wavelet and TSK fuzzy rules for stock price forecasting. IEEE Transactions On Systems, Man And Cybernetics Part C: Applications And

Reviews, 38, 802–815. http://dx.doi.org/10.1109/TSMCC. 2008.2001694.
[38]. Chang, P. C., & Liu, C. H. (2008). A TSK type fuzzy rule based system for stock price prediction. Expert Systems With Applications, 34(1), 135–144. http://dx.doi.org/10. 1016/j.eswa.2006.08.020.
[39]. Chang, P. C., Liu, C. H., Lin, J. L., Fan, C. Y., & Ng, C. S. (2009). A neural network with a case based dynamic window for stock trading prediction. Expert Systems With Applications, 36(3 PART 2), 6889–6898. http://dx.doi.org/10.1016/j.eswa.2008.08. 077.
[40]. Chang, P. C., & Wu, J. L. (2015). A critical feature extraction by kernel PCA in stock trading model. Soft Computing, 19(5), 1393–1408.
[41]. Chang, P. C., Wu, J. L., & Lin, J. J. (2016). A Takagi-Sugeno fuzzy model combined with a support vector regression for stock trading forecasting. Applied Soft Computing, 38, 831–842. http://dx.doi.org/10.1016/j.asoc.2015.10.030.
[42]. Chavarnakul, T., & Enke, D. (2018). Aintelligent technical analysis based equivolume charting for stock trading using neural networks. Expert Systems With Applications, 34, 1004–1017.
[43]. Chen, M. Y., & Chen, B. T. (2015). A hybrid fuzzy time series model based on granular computing for stock price forecasting. Information Sciences, 294, 227–241. http://dx.doi.org/10.1016/j.ins.2014.09.038.
[44]. Chen, Y. J., Chen, Y. M., & Lu, C. L. (2017). Enhancement of stock market forecasting using an improved fundamental analysis-based approach. Soft Computing, 21(13), 3735–3757. http://dx.doi.org/10.1007/s00500-016-2028-y.
[45]. Chen, Y. S., Cheng, C. H., & Tsai, W. L. (2014). Modeling fitting-function-based fuzzy time series patterns for evolving stock index forecasting. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex ProblemSolving Technologies, 41(2), 327–347. http://dx.doi.org/10.1007/s10489-014-0520- 6.
[46]. Chen, Y., & Hao, Y. (2017). A feature weighted support vector machine and Knearest neighbor algorithm for stock market indices prediction. Expert Systems With Applications, 80, 340–355.
[47]. Chen, Y., & Hao, Y. (2018). Integrating principle component analysis and weighted support vector machine for stock trading signals prediction. Neurocomputing, 321, 381–402. http://dx.doi.org/10.1016/j.neucom.2018.08.077.
[48]. Chen, A. S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan stock index. Computers And Operations Research, 30(6), 901–923. http://dx.doi.org/10.1016/ S0305-0548(02)00037-0.
[49]. Chen, Y., Lin, W., & Wang, J. Z. (2019). A dual-attention-based stock price trend prediction model with dual features. IEEE Access, 7, 148047–148058.
[50]. Chiang, W. C., Enke, D., Wu, T., & Wang, R. (2016). An adaptive stock index trading decision support system. Expert Systems With Applications, 59, 195–207. http://dx.doi.org/10.1016/j.eswa.2016.04.025.
[51]. Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems With Applications, 83, 187–205. http://dx.doi.org/10.1016/j.eswa.2017.04. 030.
[52]. Chou, J., & Nguyen, T. (2018). Forward forecast of stock price using slidingwindow metaheuristic-optimized machine-learning regression. IEEE Transactions On Industrial Informatics, 14(7), 3132–3142.
[53]. Chung, H., & Shin, K. S. (2018). Genetic algorithm-optimized long short-term memory network for stock market prediction. Sustainability, 10(10), http://dx.doi.org/10. 3390/su10103765.
[54]. Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. The Journal Of Finance, 53, 1839–1885.

[55]. Das, S. P., & Padhy, S. (2018). A novel hybrid model using teaching–learningbased optimization and a support vector machine for commodity futures index forecasting. International Journal Of Machine Learning And Cybernetics, 9(1), 97–111. http://dx.doi.org/10.1007/s13042-015-0359-0.

[56]. de Faria, E. L., Albuquerque, M. P., Gonzalez, J. L., Cavalcante, J. T., & Albuquerque, M. P. (2009). Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. Expert Systems With Applications, 36(10), 12506–12509. http://dx.doi.org/10.1016/j.eswa.2009.04.032.

[57]. Ebadati, O. M., & Mortazavi, M. T. (2018). An efficient hybrid machine learning method for time series stock market forecasting. Neural Network World, 28(1), 41–55. http://dx.doi.org/10.14311/NNW.2018.28.003.

[58]. Ebrahimpour, R., Kabir, E., & Yousef, M. R. (2007). Face detection using mixture of MLP experts. Neural Processing Letters, 26, 69–82. Ebrahimpour, R., Nikoo, H., Masoudnia, S., Yousefi, M. R., & Ghaemi, M. S. (2011). Mixture of mlp-experts for trend forecasting of time series: A case study of the tehran stock exchange. International Journal Of Forecasting, 27(3), 804–816. http://dx.doi.org/10.1016/j.ijforecast.2010.02.015.

[59]. Efendi, R., Arbaiy, N., & Deris, M. M. (2018). A new procedure in stock market forecasting based on fuzzy random auto-regression time series model. Information Sciences, 441, 113–132. E

[60]. nke, D., Grauer, M., & Mehdiyev, N. (2011). Stock market prediction with multiple regression, fuzzy type-2 clustering and neural networks. Procedia Computer Science, 6, 201–206. http://dx.doi.org/10.1016/j.procs.2011.08.038.

[61]. Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. Expert Systems With Applications, 29(4), 927–940.

[62]. Fadlalla, A., & Amani, F. (2014). Predicting next trading day closing price of Qatar exchange index using technical indicators and artificial neural networks. Intelligent Systems In Accounting, Finance And Management, 21, 209–223.

[63]. Fama, E. F. (1965). The behavior of stock market prices. The Journal Of Business, 38, 34–105.

[64]. Fama, E. (1970). Efficient capital markets: A review of the theory. The Journal Of Finance, 25, 383–417.

[65]. Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. Journal Of Financial Economics, 49, 283–306.

[66]. Farias Nazário, R. T., e Silva, J. L., Sobreiro, V. A., & Kimura, H. (2017). A literature review of technical analysis on stock markets. Quarterly Review Of Economics And Finance, 66, 115–126. http://dx.doi.org/10.1016/j.qref.2017.01.014.

[67]. Feuerriegel, S., & Gordon, J. (2018). Long-term stock index forecasting based on text mining of regulatory disclosures. Decision Support Systems, 112(June), 88–97. http://dx.doi.org/10.1016/j.dss.2018.06.008.

[68]. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal Of Operational Research, 270, 654–669.

[69]. Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. Computer Science Review, 34, Article 100190. http://dx.doi. org/10.1016/j.cosrev.2019.08.001.

[70]. Göçken, M., Özçalıcı, M., Boru, A., & Dosdoğru, A. T. (2019). Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection. Neural Computing And Applications, 31(2), 577–592.

[71]. Gocken, M., ÖZÇALICI, M., Boru, A., & Dosdoğru, A. T. (2016). Integrating metaheuristics and artificial neural networks for improved stock price prediction. Expert Systems With Applications, 44, 320–331.

[72]. Gorenc Novak, M., & Velušček, D. (2016). Prediction of stock price movement based on daily high prices. Quantitative Finance, 16(5), 793–826. http://dx.doi.org/10.1080/14697688.2015.1070960.

[73]. Gowthul Alam, M. M., & Baulkani, S. (2019). Local and global characteristics-based kernel hybridization to increase optimal support vector machine performance for stock market prediction. Knowledge And Information Systems, 60(2), 971–1000. http://dx.doi.org/10.1007/s10115-018-1263-1.

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

[75]. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal Of Machine Learning Research, 3, 1157–1182.

[76]. Hadavandi, E., Shavandi, H., & Ghanbari, A. (2010). Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. Knowledge-Based Systems, 23(8), 800–808. http://dx.doi.org/10.1016/j.knosys.2010.05.004.

[77]. Hassan, M. R., Nath, B., & Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. Expert Systems With Applications, 33(1), 171–180.

[78]. Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. Expert Systems With Applications, 124, 226–251.

[79]. Horata, P., Chiewchanwattana, S., & Sunat, K. J. N. (2013). Robust extreme learning machine. Neurocomputing, 102, 31–44.

[80]. Hoseinzade, E., & Haratizadeh, S. (2019). CNNPred: CNN-based stock market prediction using a diverse set of variables. Expert Systems With Applications, 129, 273–285. http://dx.doi.org/10.1016/j.eswa.2019.03.029.

[81]. Hsieh, L. F., Hsieh, S. C., & Tai, P. H. (2011). Enhanced stock price variation prediction via DOE and BPNN-based optimization. Expert Systems With Applications, 38(11), 14178–14184. http://dx.doi.org/10.1016/j.eswa.2011.04.229.

[82]. Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with Google Trends. Neurocomputing, 285, 188–195. http://dx.doi.org/10.1016/j.neucom.2018.01.038.

[83]. Huang, J. Y., & Liu, J. H. (2020). Using social media data mining technology to improve stock price forecast accuracy. Journal Of Forecasting, 39, 104–116.

[84]. Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers And Operations Research, 32, 2513–2522.

[85]. Huang, C. L., & Tsai, C. Y. (2009). A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. Expert Systems With Applications, 36, 1529–1539.

[86]. Hyup Roh, T. (2007). Forecasting the volatility of stock price index. Expert Systems With Applications, 33(4), 916–922. http://dx.doi.org/10.1016/j.eswa.2006.08.001.

[87]. Javedani Sadaei, H., & Lee, M. H. (2014). Multilayer stock forecasting model using fuzzy time series. The Scientific World Journal, 2014, http://dx.doi.org/10.1155/2014/610594.

[88]. Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress. Expert Systems With Applications, 184, Article 115537. http://dx.doi.org/10.1016/j.eswa.2021.115537.

[89]. Jiang, M., Liu, J., Zhang, L., & Liu, C. (2020). An improved stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. Physica A: Statistical Mechanics And Its Applications, 541(258), Article 122272. http://dx.doi.org/10.1016/j.physa.2019.122272.

[90]. Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. Expert Systems With Applications, 178, Article 115019.

[91]. Kao, L. J., Chiu, C. C., Lu, C. J., & Chang, C. H. (2013). A hybrid approach by integrating wavelet-based feature extraction with MARS

and SVR for stock index forecasting. Decision Support Systems, 54(3), 1228–1244.

[92]. Kara, Y., Acar Boyacioglu, M., & Baykan, O. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul stock exchange. Expert Systems With Applications, 38, 5311–5319.

[93]. Kaur, G., Dhar, J., & Guha, R. K. (2016). Minimal variability OWA operator combining ANFIS and fuzzy c-means for forecasting BSE index. Mathematics And Computers In Simulation, 122, 69–80. http://dx.doi.org/10.1016/j.matcom.2015.12.001.

[94]. Kazem, A., Sharifi, E., Hussain, F. K., Saberi, M., & Hussain, O. K. (2013). Support vector regression with chaos-based firefly algorithm for stock market price forecasting. Applied Soft Computing, 13(2), 947–958.

[95]. Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. Applied Mathematical Finance, 1–20.

[96]. Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55, 307–319.

[97]. Kim, K.-j., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert Systems With Applications, 19, 125–132.

[98]. Kim, S., Ku, S., Chang, W., Chang, W., Chang, W., & Song, J. W. (2020). Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques. IEEE Access, 8, 111660–111682. http://dx.doi.org/10.1109/ACCESS.2020.3002174.

[99]. Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele University, UK and National ICT Australia.

[100]. Kitchenham, B., & Brereton, P. (2013). A systematic review of systematic review process research in software engineering. Information And Software Technology, 55, 2049–2075.

[101]. Klemm, P., Bunnell, D., Cullen, M., Soneji, R., Gibbons, P., & Holecek, A. (2003). Online cancer support groups: a review of the research literature.. Computers, Informatics, Nursing : CIN, 21, 136–142. http://dx.doi.org/10.1097/00024665-200305000-00010.

[102]. Kubat, M. (2017). An introduction to machine learning (2nd Ed.). Springer Publishing Company, Incorporated.

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

[104]. Laboissiere, L. A., Fernandes, R. A. S., & Lage, G. G. (2015). Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. Applied Soft Computing, 35, 66–74. http://dx.doi.org/10.1016/j.asoc.2015.06.005.

[105]. Lai, R. K., Fan, C. Y., Huang, W. H., & Chang, P. C. (2009). Evolving and clustering fuzzy decision tree for financial time series data forecasting. Expert Systems With Applications, 36(2 PART 2), 3761–3773. http://dx.doi.org/10.1016/j.eswa.2008.02.025.

[106]. Lam, M. (2004). Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis. Decision Support Systems, 37(4), 567–581. http://dx.doi.org/10.1016/S0167-9236(03)00088-5.

[107]. Lee, S. W., & Kim, H. Y. (2020). Stock market forecasting with super-high dimensional time-series data using ConvLSTM, trend sampling, and specialized data augmentation. Expert Systems With Applications, 161, Article 113704.

[108]. Leigh, W., Purvis, R., & Ragusa, J. M. (2002). Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support. Decision Support Systems, 32, 361–377.

[109]. Leung, M. T., Daouk, H., & Chen, A. S. (2000). Forecasting stock indices: A comparison of classification and level estimation models. International Journal Of Forecasting, 16(2), 173–190. http://dx.doi.org/10.1016/S0169-2070(99)00048-5.

[110]. Li, Y., Bu, H., Li, J., & Wu, J. (2020). The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning. International Journal Of Forecasting, 36(4), 1541–1562. http://dx.doi.org/10.1016/j.ijforecast. 2020.05.001.

[111]. Li, B., Chan, K., Ou, C. X., & Sun, R. (2017). Discovering public sentiment in social media for predicting stock movement of publicly listed companies. Information Systems, 69, 81–92.

[112]. Liao, Z., & Wang, J. (2010). Forecasting model of global stock index by stochastic time effective neural network. Expert Systems With Applications, 37(1), 834–841. http://dx.doi.org/10.1016/j.eswa.2009.05.086.

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

[114]. Lim, K. P., & Brooks, R. (2011). The evolution of stock market efficiency over time: a survey of the empirical literature. Journal Of Economic Surveys, 25, 69–108.

[115]. Liu, H., & Long, Z. (2020). An improved deep learning model for predicting stock market price time series. Digital Signal Processing: A Review Journal, 102, Article 102741. http://dx.doi.org/10.1016/j.dsp.2020.102741.

[116]. Lo, A. W. (2004). The adaptive markets hypothesis. The Journal Of Portfolio Management, 30, 15–129.

[117]. Lohrmann, C., & Luukka, P. (2019). Classification of intraday S&P500 returns with a random forest. International Journal Of Forecasting, 35(1), 390–407. http://dx.doi. org/10.1016/j.ijforecast.2018.08.004.

[118]. Lohrmann, C., Luukka, P., Jablonska-Sabuka, M., & Kauranne, T. (2018). A combination of fuzzy similarity measures and fuzzy entropy measures for supervised feature selection. Expert Systems With Applications, 110, 216–236.

[119]. Lu, C. J. (2013). Hybridizing nonlinear independent component analysis and support vector regression with particle swarm optimization for stock index forecasting. Neural Computing And Applications, 23(7–8), 2417–2427. http://dx.doi.org/10.1007/ s00521-012-1198-5.

[120]. Lu, C. J., Lee, T. S., & Chiu, C. C. (2009). Financial time series forecasting using independent component analysis and support vector regression. Decision Support Systems, 47(2), 115–125.

[121]. Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. Complexity, 2020, http://dx.doi.org/10.1155/2020/6622927.

[122]. Lu, C. J., & Wu, J. Y. (2011). An efficient CMAC neural network for stock index forecasting. Expert Systems With Applications, 38(12), 15194–15201. http://dx.doi. org/10.1016/j.eswa.2011.05.082.

[123]. Maknickiene, N., & Maknickas, A. (2013). Financial market prediction system with Evolino neural network and delphi method. Journal Of Business Economics And Management, 14(2), 403–413.

[124]. Malagrino, L. S., Roman, N. T., & Monteiro, A. M. (2018). Forecasting stock market index daily direction: A Bayesian network approach. Expert Systems With Applications, 105, 11–22. http://dx.doi.org/10.1016/j.eswa.2018.03.039.

[125]. Malkiel, B., & Mullainathan, S. (2005). Market efficiency versus behavioral finance. Journal Of Applied Corporate Finance, 17, 124–136.

[126]. Mo, H., & Wang, J. (2018). Return scaling cross-correlation forecasting by stochastic time strength neural network in financial market dynamics. Soft Computing, 22(9), 3097–3109. http://dx.doi.org/10.1007/s00500-017-2564-0.

[127]. Mostafa, M. M. (2010). Forecasting stock exchange movements using neural networks: Empirical evidence from Kuwait. Expert Systems With Applications, 37(9), 6302–6309. http://dx.doi.org/10.1016/j.eswa.2010.02.091.

[128]. Murphy, J. J. (1999). Technical analysis of the financial markets: a comprehensive guide

to trading methods and applications, Vol. 2. Prentice Hall Press.

[129]. Na, S. H., & Sohn, S. Y. (2011). Forecasting changes in Korea composite stock price index (KOSPI) using association rules. Expert Systems With Applications, 38(7), 9046–9049. http://dx.doi.org/10.1016/j.eswa.2011.01.025.

[130]. Nermend, Y., & Alsakaa, K. (2017). Back-propagation artificial neural networks in stock market forecasting . An application to the warsaw stock exchange WIG20. The IEB International Journal Of Finance, 15, 88–99.

[131]. Niaki, S. T. A., & Hoseinzade, S. (2013). Forecasting S&P 500 index using artificial neural networks and design of experiments. Journal Of Industrial Engineering International, 9(1), 1–9. http://dx.doi.org/10.1186/2251-712X-9-1.

[132]. O'Connor, N., & Madden, M. G. (2006). A neural network approach to predicting stock exchange movements using external factors. Knowledge-Based Systems, 19, 371–378.

[133]. Olson, D., & Mossman, C. (2003). Neural network forecasts of Canadian stock returns using accounting ratios. International Journal Of Forecasting, 19(3), 453–465. http: //dx.doi.org/10.1016/S0169-2070(02)00058-4.

[134]. Pai, P. F., & Lin, C. S. (2005). ARIMA,artificial neural networks,stock prices,support vector machines,time series forecasting. Omega, 33, 497–505.