Expert Systems With Applications 156 (2020) 113464

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

Stock market movement forecast: A Systematic review☆

O Bustos[a,1,*], A. Pomares-Quimbaya[a,2]

Facultad de Ingeniería. Departamento de Ingeniería de Sistemas, Pontificia Universidad Javeriana, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 24 October 2019
Revised 21 February 2020
Accepted 16 April 2020
Available online 20 April 2020

Keywords:
Stock market forecast
Machine learning
Financial modeling

ABSTRACT

Achieving accurate stock market models can provide investors with tools for making better data-based decisions. These models can help traders to reduce investment risk and select the most profitable stocks. Furthermore, creating advanced models enable the usage of non-traditional data like historical stock prices and news. There are several review articles about financial problems, including stock market analysis and forecast, currency exchange forecast, optimal portfolio selection, among others. However, the recent advances in machine learning techniques, like Deep Learning, Text Mining Techniques, and Ensemble Techniques, raises the need to perform an updated review. This study aims to fill this gap by providing an updated systematic review of the forecasting techniques used in the stock market, including their classification, characterization and comparison. The review is focused on studies on stock market movement prediction from 2014 to 2018, obtained from the scientific databases Scopus and Web of Science. Besides, it analyzes surveys and other reviews of recent studies published in the same time frame and the same databases.

1. Introduction

The stock market forecast has been a challenging problem to solve. The efficient-market hypothesis presented by Fama (1995) suggests that, in the efficient information markets, stock prices behave like a random walk and it is impossible to forecast direction and magnitude changes. He proposed three categories of efficiency: weak form, where past price movements can't be used to predict the future ones; semi-strong form, where neither the past price movements and any public information is relevant for predicting the market; strong-form, where none of the information, public or private can be used to forecast the market.

Despite Fama's hypothesis, the scientific community has proposed different ways to forecast stock market (Cavalcante, Brasileiro, Souza, Nobrega, & Oliveira, 2016). The first one is the fundamental analysis, where underlying factors that affect companies or industries are used as predictive attributes. The second one is the technical analysis, where the predictive attributes are mainly historical prices and volumes.

Technical Analysis is the most common approach in the literature (Cavalcante et al., 2016; George Atsalakis, 2009; Rubén Aguilar-Rivera, 2015), using as input stock prices or indicators derived from it. Technician analysts argue that all new information, like news and macroeconomic variables, are already represented in stock prices. Therefore it is enough to analyze the patterns of the price trends to predict the stock market. Technical indicators have been extensively studied and have been used as stock signals for indicating when to buy or sell a stock, as presented in the article by Nazário, e Silva, Sobreiro, and Kimura (2017), where they survey studies of technical analysis on the stock market throughout 55 years. However, some studies like the presented in Park and Irwin's survey (Park & Irwin, 2007) showed that trading strategies based on technical indicators have limited results.

Fundamental analysis is less common in literature because it is harder to build models that understand why a stock is fluctuating. The most common information used is related to macroeconomic time series, like Gross Domestic Product, interest rates, currency exchange rates, customer price index, among others (Boyacioglu & Avci, 2010). Other sources of information are general as financial news, but its unstructured nature and non-continuous behavior make them harder to use. Text mining techniques have been applied to handle this complexity. Most recently, social network analysis has proved useful for stock forecasting (Bollen, Mao, & Zeng, 2011), using sentiment indexes and other derived series as inputs.

☆ This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author.
E-mail addresses: oscar.bustos@javeriana.edu.co (O. Bustos), pomares@javeriana.edu (A. Pomares-Quimbaya).
[1] PhD. Student
[2] Full Professor, Tel: +57-1-3208320

https://doi.org/10.1016/j.eswa.2020.113464
0957-4174/

This paper presents an updated systematic review of the state of the art in the stock market forecast considering fundamental and technical analysis from 2014 to 2018. The review describes the modeling techniques, the kind of data set, and the performance indicators used in the revised articles. In order to compare the recent works, this review focuses on the forecasting problem as a classification problem, where the models predict the direction of the movement of the stock price, like up/down. In addition, it includes a discussion on the new challenges in the stock market forecast.

The first part of the article describes the systematic review methodology and summarizes other financial relevant surveys. The second part of the review presents the scientific works, describing its inputs and outputs. In the third part, the articles are classified and detailed depending on the modeling techniques used, like bio-inspired, linear classifiers, and hybrid ones. Finally, the article discusses the main challenges, conclusions and future work recommendations.

2. Research methodology and related works

As a first step of the systematic review, survey articles focused on the prediction of the stock market were reviewed. In order to perform the search, two of the primary scientific paper databases, Web of Science and Scopus, were selected. Only the articles with the words "survey" or "review" as keywords or part of the abstract were taken into account, to discard non-survey articles. Another filter criteria were if the words "forecast" or "prediction" and "stock market" or "financial market" were present in the abstract or keywords. Only survey articles that studied the financial market itself were selected. Fourteen articles were found, which were characterized by the number of references, the number of citations, and the time range in which the review was expanded.

Table 1 summarizes the surveys found in the literature. Columns indicate the time frame analyzed by the review, the number of studies included, the number of citations it has, the prediction problems on which the review focuses, and the techniques analyzed. In general, these articles have been beneficial for the scientific community, because the number of references to these works is more than 1000. Most of these works focus in Stock Market (SM), but some of them describe other financial problems like Exchange Rates forecasting (ER), Financial Distress analysis (FD), Credit Scoring (CS), and Interest Rate forecasting (IR). Furthermore, some of the works survey most of the relevant machine learning techniques (ML), while others focused in a specific technique like the Artificial Neural Networks (ANN), Support Vector Machines (SVM), Text Mining (TM), and Bio-Inspired Computing (BI).

As presented in Table 1 most of the recent reviews are focused on text mining (Khadjeh Nassirtoussi, Aghabozorgi, Ying Wah, & Ngo, 2014; Nardo, Petracco-Giudici, & Naltsidis, 2016; Xing, Cambria, & Welsch, 2018). Given the last advances in social network analysis, it has been possible to include this kind of information as a predictive input to the forecasting models. Besides, text processing techniques have evolved, and now those allow extracting more complex information about text bodies. Xing et al. (2018) selected the most important articles related to natural language-based financial forecasting. They listed the types of financial texts used as predictive input and how they are processed. They also described the algorithms involved in the modeling and implementation details. Khadjeh Nassirtoussi et al. (2014) reviewed Text mining for market prediction. They classified articles based on the nature of textual input, like Financial news or Tweets, and the class of the market, the preprocessing procedure and types of modeling techniques. Nardo et al. (2016) searched for works in the literature that links changes in the stock returns and the internet related information. They showed the limitation of those works and proposed some modeling techniques for future research, such as herding behavior and contagions, in order to understand the relation with the web buzz prediction power.

Some of the review works compare most of the relevant forecasting techniques (Atsalakis & Valavanis, 2009; Bahrammirzaee, 2010; Cavalcante et al., 2016; Preethi & Santhi, 2012; Yoo, Kim, & Jan, 2005). One of the most recent reviews in the field is the article by Cavalcante et al. (2016). They analyze not only forecasting algorithms but also all variety of algorithms related to other problems in financial markets. Among those problems, they present feature selection, clustering, segmentation, and outlier detection. They also propose an architecture for autonomous trading in real markets. Although this is a complete review of the computational application in financial markets, this survey expands the revision on the stock market prediction from several types of inputs and algorithms. Atsalakis and Valavanis (2009) describe more than 100 articles using the following characteristics: predictive variables, modeling techniques, benchmarks, and performance measures. Their focus is on neural networks and fuzzy networks, where they describe the type of transfer function, the type of membership function, the network architecture, and the training method. This review inspired the structure to be followed in the survey, even though the present article covered other computational learning techniques not worked in Atsalakis and Valavanis (2009) revision. Besides, it is important to update this review because, after 2009, there has been produced a significant amount of new research work on stock market prediction using techniques from the family of neural networks.

Table 1
List of review articles. Stock Market (SM), Exchange Rates forecasting (ER), Financial Distress analysis (FD), Credit Scoring (CS), Interest Rate forecasting (IR), Machine Learning techniques (ML), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Text Mining (TM),Bio-Inspired Computing (BI).

Author	Year	Start Date	End Date	Citations	References	Problem	Technique
Tkáč and Verner (2016)	2016	1994	2015	29	436	SM,ER,FD,CS,IR	ANN
Bahrammirzaee (2010)	2010	1979	2010	109	281	SM,CS	ML
Cavalcante et al. (2016)	2016	2009	2015	43	132	SM,ER,FD,CS	ML
Xing et al. (2018)	2018	1998	2016	7	128	SM	TM
Khadjeh Nassirtoussi et al. (2014)	2014	1996	2014	107	117	SM	TM
Sapankevych and Sankar (2009)	2009	1997	2009	332	105	SM	SVM
Atsalakis and Valavanis (2009)	2009	1992	2006	284	100	SM	ML
Nardo et al. (2016)	2016	1998	2014	10	87	SM	TM
Vui et al. (2013)	2013	1988	2013	20	51	SM	ANN
Yoo et al. (2005)	2005	1991	2004	24	48	SM	ML
Krollner et al. (2010)	2010	2000	2009	18	46	SM,ER,IR	ML
Li and Ma (2010)	2010	1990	2002	31	40	SM,CS,FD	ANN
Ripon and Rajon (2016)	2016	2000	2015	2	24	SM	BI
Preethi and Santhi (2012)	2012	2005	2011	14	20	SM	ML

At the same time, there are many review works focused in a specific technique for stock market forecast (Krollner, Vanstone, & Finnie, 2010; Li & Ma, 2010; Sapankevych & Sankar, 2009; Tkáč & Verner, 2016; Vui, Soon, On, Alfred, & Anthony, 2013). Techniques such as SVM and ANN have been worked extensively by the scientific community. These techniques have shown great predictive power, non-linear modeling behavior, and robustness to overfitting. Tkáč and Verner (2016) described the application of ANN in a business application, where the most important research aimed to solve financial distress and bankruptcy problems, stock price forecasting, and decision support. They described the most common Neural Networks hyper-parameters, like the kind of neural network and the type of learning algorithms. They also analyzed some bibliometric variables like the number of citations, the most occurring journals, and the most cited papers. Vui et al. (2013) proposed some guidelines and steps when forecasting is performed using ANN.

Once analyzed these reviews, it can be concluded that there are vast amounts of work that are being published each year. Tkáč and Verner (2016) surveyed more than 400 articles and Bahrammirzaee (2010) more than 200. Most of the other reviews referenced more than 100 articles. Those surveys are quickly outdated, given the considerable amount of new works published each year.

Furthermore, some of the reviews focused on a specific set of techniques, making it harder to compare the most performing works. For this reason, it is crucial to perform an updated systematic review with the latest trends that include both text mining and the latest machine learning techniques.

Once the review of similar works was finished, it was defined the search equation to find state of the art in Stock Market Forecasting. Web of Science and Scopus databases were surveyed to update the review. The search period selected was between 2014 and 2018.

This survey aims to answer the following research questions: 1) What are the different phases of stock market forecasting? Here the research aims to describe and compare the steps needed to perform the stock market forecasting. For each step, we describe the different approaches reported in the articles and conclude their advantages or disadvantages. 2) How successful have been computational algorithms in forecasting stock market prices? The primary interest is to compare the accuracy of the models given the modeling technique.

The selection of the studies was made considering the following inclusion criteria:

IC1: The primary purpose is on stock market direction forecasting
IC2: The study reports evaluation metrics

Said that, the search query included two groups of keywords: Group 1: “Financial Price Model”, “Financial Markets”, “Shares Market”, “Stock Market”, “Financial Volatility”; Group2: “Automated Trading”, “Stock Return”, “Forecasting”, “Predicting”, “Predictive Model”, “Forecast Performance”, “Price Models”, “Algorithm”, “Machine Learning”, “Computational Intelligence”, “Big Data”, “Time Series Analysis”. The OR operator was used for the keywords in the same group, and the AND operator was used between the groups.

The selection process is presented in Fig. 1. After running the search against the databases, 219 articles matched the search equation. From that search, it was excluded 25 duplicated articles. After reading the titles and their abstract, other 101 articles were excluded from this review because their primary purpose is not stock market direction forecasting. After making a full reading of those articles, it was discarded 43 works because did not report an evaluation metric. Experts suggested the inclusion of 3 relevant articles to be included in the review. Finally, the review was carried out among 53 articles.

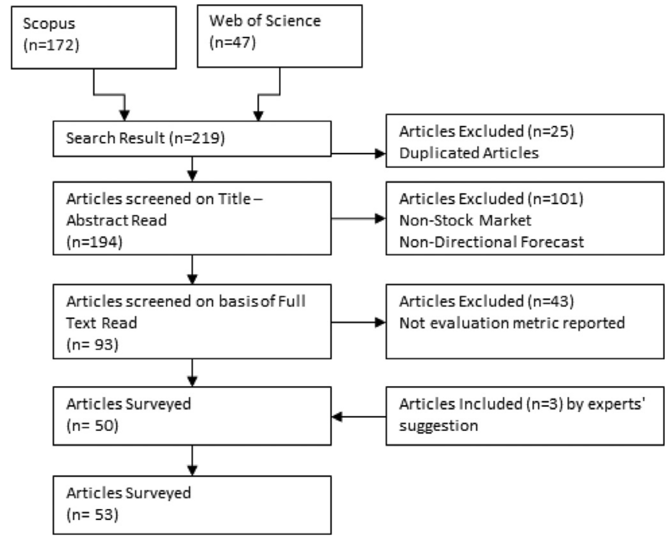

Fig. 1. Flow diagram of the selected articles.

Table 2
Ranking of main Journals .

Journal	Articles count
Expert Systems with Applications	8
Neurocomputing	3
Lecture Notes in Computer Science	2
Information Systems	1
Big Data Research	1
International Journal of Machine Learning and Computing	1
Computational Intelligence and Neuroscience	1
IEEE Transactions on Computational Social Systems	1
Economic Computation & Economic Cybernetics Studies & Research	1
International Journal of Circuits, Systems and Signal Processing	1
European Journal of Operational Research	1
Communications in Computer and Information Science	1
Knowledge-Based Systems	1
Mathematical Problems in Engineering	1
Advances in Intelligent Systems and Computing	1
PLoS ONE	1
ACM Transactions on Information Systems	1
World Wide Web	1
IEEE Systems Journal	1

Most of the papers reviewed are from indexed journals, as presented in Table 2. The most relevant journals are Expert Systems with Applications with 8 articles, Neurocomputing with 3 articles, and Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) with 2 articles surveyed. The other articles reviewed were presented at international conferences, almost entirely published by the Institute of Electrical and Electronics Engineers Inc.

As can be seen in Fig. 2, most of the selected works were written in 2018. These new works are summarized in Table 3 and described in the following sections. Stock/index indicates if the proposed model is focused on predicting a stock, an index or both. The country, as well as the time frame is also described. And finally, the type of input and output involved in each study was detailed (Fig. 3).

3. Data sourcing

The work done on stock market prediction can be classified depending on the type of inputs they use. A large part of the articles

Table 3
Stock market prediction projects.

Author	Stock/Index	Country	Initial Year	Final Year	Type of Input
Arévalo, García, Guijarro, and Peris (2017)	Index	USA	2000	2013	Technical Indicator
Bustos, Pomares, and Gonzalez (2018)	Stocks	Colombia	2010	2017	Technical Indicator
Cen, Ruta, and Ruta (2017)	Stocks	Multiple countries			Blogs
Chai et al. (2015)	Index	China	2009	2012	Economic Indicators
Chakraborty et al. (2018)	Both	USA	2016	2016	Social Network
Coyne et al. (2018)	Stocks	USA	2016	2017	Social Network
Dang and Duong (2016)	Index	Vietnam	2014	2015	Social Network
Dash and Dash (2016)	Index	Multiple countries	2010	2014	Technical Indicator
Dingli and Fournier (2017)	Index	USA	2003	2016	Market Information, Technical Indicator, Economic Indicators
Di Persio and Honchar (2016)	Index	USA	1950	2016	Market Information
Feuerriegel and Fehrer (2016)	Stocks	Germany	2004	2011	News
Fischer and Krauss (2018)	Index	USA	1992	2015	Market Information
Ghanavati et al. (2016)	Index	China	2014	2015	Technical Indicator, News
Gonzalez et al. (2015)	Index	Multiple countries	1989	1998	Market Information, Technical Indicator, Economic Indicators
Gunduz et al. (2017)	Stocks	Turkey	2011	2015	Technical Indicator
Guo et al. (2017)	Index	China	2000	2004	Technical Indicator
Hu et al. (2018)	Index	USA	2010	2015	Market Information, Social Network
Huang and Li (2017)	Index	China	2015	2016	Market Information
Huang et al. (2018)	Stocks	USA	2008	2009	Market Information, Technical Indicator
Ince and Trafalis (2017)	Stocks	USA	2007	2015	Market Information, Technical Indicator
Kamble (2018)	Stocks	Multiple countries	2012	2017	Market Information, Economic Indicators
Kia, Haratizadeh, and Shouraki (2018)	Index	Multiple countries			
Kim and Enke (2016)	Index	Korea	2007	2014	Technical Indicator
Labiad et al. (2016)	Stocks	Morocco	2008	2015	Technical Indicator
Leito, Neves, and Horta (2016)	Index	USA	2011	2014	Technical Indicator
Li, Chen, et al. (2016)	Stocks	China	2011	2011	Market Information, News
Li, Tam, and Yeung (2016b)	Index	China	2011	2013	Technical Indicator
Li et al. (2017)	Stocks	USA	2011	2012	Social Network
Liu et al. (2018)	Stocks	USA	2011	2017	Technical Indicator, News
Malagrino, Roman, and Monteiro (2018)	Index	Brazil	2005	2012	Market Information
McCluskey and Liu (2017)	Index	USA	2000	2012	Market Information, Economic Indicators
Mingyue et al. (2016)	Index	Japan	2013	2014	Technical Indicator
Nelson, Pereira, and De Oliveira (2017)	Index	Brazil	2008	2015	Technical Indicator
Pagolu et al. (2017)	Stocks	USA	2015	2016	Social Network
Patel et al. (2015)	Stocks	India	2003	2012	Technical Indicator
Porshnev et al. (2015)	Index	USA	2013	2014	Social Network
Qiu and Song (2016)	Index	Japan	2007	2013	Technical Indicator
Ren, Wu, and Liu (2018)	Index	China	2014	2016	Blogs
Shynkevich, McGinnity, Coleman, Li, and Belatreche (2014)	Stocks	USA	2000	2011	Technical Indicator
Shynkevich et al. (2015)	Stocks	USA	2009	2014	News
Sun et al. (2017)	Stocks	China	2016	2016	Social Network
Tsantekidis et al. (2017)	Stocks	Finland	2010	2010	Market Information
Tüfekci (2016)	Index	Turkey	2007	2012	Economic Indicators
Van Den Poel et al. (2016)	Stocks	USA	2014	2015	Market Information, Technical Indicator
Wang, Liu, et al. (2018)	Index	USA	2012	2017	Market Information
Wang, Xu, et al. (2018)	Stocks	China	2012	2015	Technical Indicator, Social Network
Weng et al. (2017)	Stocks	USA	2012	2015	Market Information, Technical Indicator, Social Network
Xu and Kešelj (2014)	Stocks	USA	2012	2012	Market Information, Social Network
Yang et al. (2017)	Stocks	China	2014	2015	Market Information, Technical Indicator, Economic Indicators
Zhang et al. (2018)	Stocks	China	2010	2016	Technical Indicator
Zhong and Enke (2017)	Index	USA	2003	2013	Economic Indicators, Market Information
Zhou, Chan, and Ou (2018)	Stocks	USA	2000	2001	Social Network

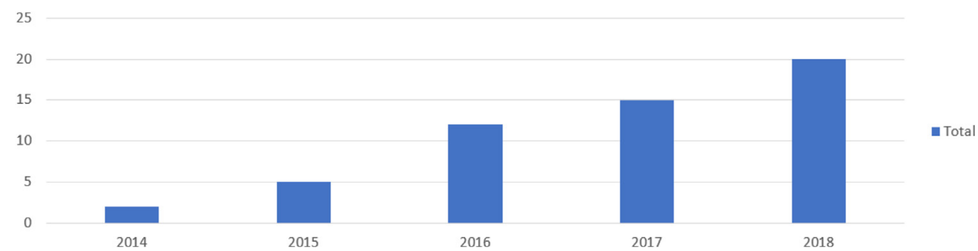

Fig. 2. Count of articles by publication year.

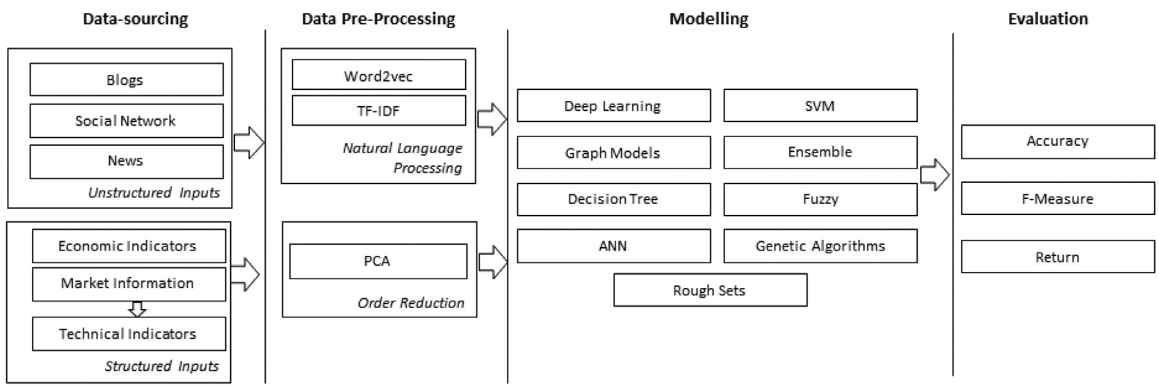

Fig. 3. Phases of the stock market modeling .

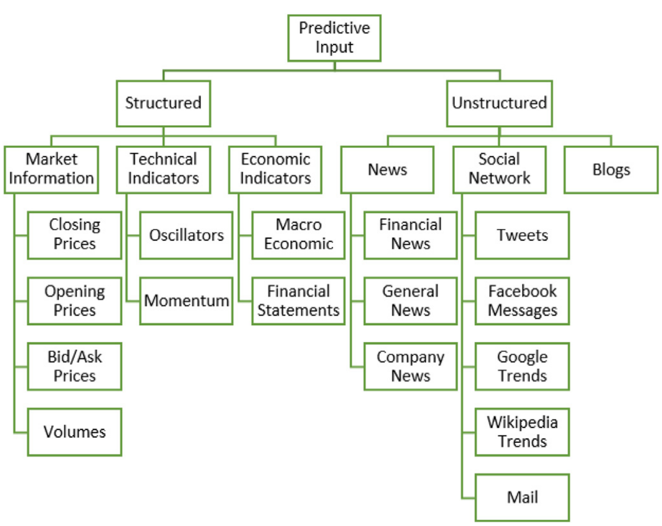

Fig. 4. Classifications of inputs.

reviewed use structured type inputs, for which processing techniques already exist, and their importance has been extensively studied. Most recent ones allow the use of unstructured information, which is more difficult to process and to extract useful information. Fig. 4 shows a proposed taxonomy for the inputs used to forecast the stock market in the analyzed studies.

3.1. Structured inputs

The structured information refers to data groups with a predefined skeleton, organized in tabular form, where the characteristics or attributes can be described as columns of a table. That structure makes information more accessible to navigate, and simple or complex searches can be done without further effort. Most articles use this type of information, which is usually open and exposed through API programming interfaces. The most common is the time series of historical stock prices, which can be used directly by different computational models.

3.1.1. Stock values

Given the technical analysis approach, stock prices reflect all the information required to understand market behavior. In this way, the important thing is to analyze the series of time corresponding to the prices. Generally, this information is public and free and can be downloaded from the pages of the stock markets (such as Nasdaq Kazem, Sharifi, Hussain, Saberi, & Hussain (2013)), third parties (such as Yahoo Finance Wen, Yang, Song, & Jia (2010)). Besides, some companies like Bloomberg (Ding, Zhang, Liu, & Duan, 2015) provide paid services with more information related to stock prices.

In some articles, daily stock information is used, which consists of the opening price (OP), closing price (CP), the maximum (MAX) and minimum price (MIN), and the volume (VOL) of transactions performed Wang, Liu, Shang, and Wang (2018) Fischer and Krauss (2018) Di Persio and Honchar (2016). Closing prices are the most commonly used information, but the volume and ranges have also shown value in the prediction. Most of the studies employ a time-span of 1000 days, that can be handled easily for most of the machine learning algorithms.

In addition, there are other studies that use intraday information for prediction (Huang & Li, 2017; Tsantekidis et al., 2017). The most fine-grained intraday information is the bid-ask price for a stock. When a stock is being traded in an exchange, there are buyers and sellers interested in trading that stock. Ask price is the minimum price a seller is willing to accept, while the bid price is the maximum price that the buyer offers to pay for the share. The consolidation of all these prices leads to an enormous number of points having to be recorded to predict the intraday price.

3.1.2. Technical indicators

Technical indicators have been useful for predicting the stock market. These have been increasing in sophistication, and are already part of the language of brokers. Technical indicators can summarize the behavior or trends in the time series, making their

Table 4
Technical Indicator's mathematical equations for time tC_t is the closing price of the stock or stock index. H_n is the highest price traded in a day L_n is the lowest price traded in a day .

Indicator	Equation
SMA_t	$\frac{C_t+C_{t-1}+\cdots+C_{t-9}}{n}$
WMA_t	$\frac{(n)(C_t)+(n-1)(C_{t-1})+\cdots+C_{t-9}}{n+(n-1)+\cdots+1}$
K_t	$\frac{C_t-LL_{t-n}}{HH_{t-n}-LL_{t-n}} \times 100$
D_t	$\frac{\sum_{i=0}^{n-1} K_t}{n}$
MOM_t	$C_t + C_{t-10}$
RSI_t	$100 - \frac{100}{1+(\sum_{i=0}^{n-1} UP_{t-i}/n)/(\sum_{i=0}^{n-1} DOWN_{t-i}/n)}$
$MACD_t$	$EMA12_t - EMA26_t$
LW_t	$\frac{H_n-C_t}{H_n-L_n} \times 100$
AD_t	$\frac{H_t-C_{t-1}}{H_t-L_t}$
CCI_t	$\frac{M_t-SM_t}{0.015D_t}$

adoption more appropriate than raw prices. This representation allows applying techniques over a summarized representation instead of the stock price time series itself, simplifying the machine learning models.

There are two main categories in Technical Indicators, Trend indicators, and Oscillators. The former category is focused on identifying the movement direction of the stock value. The latter category allows identifying the turning points in the time series.

The most used trend indicators are Moving Averages and Momentum. Simple Day Moving Average (SMA) does a summary of the last day's performance. It is used with different long-term averages for uptrend forecast when the crossover of trends happens. Besides, the Weighted day moving average (WMA) is another well-known indicator that weights more recent prices than past prices. An alternative and a more straightforward trend indicator is the momentum, which measures price differences over time (Table 4). Momentum (MOM) supports the recognition of trend-lines, depending on whether it is valued above or below zero. MOM is calculated as the difference between the two SMA. Similar to MOM, the Moving Average Convergence Divergence (MACD), developed by Appel (1985), can be calculated as the difference between two exponential moving averages.

The most common oscillators are the Relative Strength Index (RSI), Commodity Channel Index (CCI), Williams R, and Stochastic Oscillators. Lane (1985) designed the Stochastic Oscillators (K and D) as a momentum indicator of a stock price. This indicator alerts when a stock is oversold or overbought. Additionally, Wilder Jr (1986) developed in 1986, the RSI indicator, which is also a momentum oscillator like K and D. This indicator takes values between 0 and 100. Additionally, Williams (1985) developed the R technical indicator, and unlike the Stochastic Oscillator K, values near 100 indicate oversold, and values near 0 indicate overbought. Accumulation/Distribution oscillator (A/D), proposed by Williams, is a momentum indicator that includes additional information from the market, live the highs, lows, and closing prices. This oscillator is used in conjunction with the share price and can predict an uptrend or downtrend. Lambert (1983) developed the technical indicator CCI in 1983. When this indicator is valued more than 100, the stock is supposed to rise.

As presented in Table 5, most of the studies used a combination of several technical indicators as predictors of the stock market. Technical indicators may complement each other to give a more accurate answer on future stock prices.

3.1.3. Macroeconomic indicators and financial reports

Fundamental analysis uses economic indicators to understand how the stock price changes are related to external and internal changes in a company. The hypothesis is that, depending on the health of a country's economy, one can estimate the growth of a company, and it could impact the stock value.

One of the most popular economic indicator used as predictive input is the exchange rates. Zhong and Enke (2017) trained their models using the prices between the USD and four other currencies. In addition to using rates comparison like EUR/USD, Dingli and Fournier (2017) used the cryptocurrencies rates, in particular, the BITCOIN/USD rates. McCluskey and Liu (2017) used EUR/USD, AUD/USD, USD/JPY and USD as predictive inputs, while Gonzalez, Padilha, and Barone (2015) used USD, EUR, and CYN rates.

Furthermore, there are selected macroeconomic indicators that have been proven to be useful, like commodities prices (Dingli & Fournier, 2017; McCluskey & Liu, 2017; Tüfekci, 2016), Consumer price index (CPI) (Tüfekci, 2016; Yang, Rao, Hong, & Ding, 2017) and Interest rates on Treasury bill (Zhong & Enke, 2017).

Table 5
Technical indicator used in predictive models. Simple Day Moving Average (SMA), Weighted day Moving Average (WMA), Momentum (MOM), Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), Commodity Channel Index (CCI), Rate Of Change (ROC), Moving average oscillators (MAO), Accumulation/Distribution Oscillator (A/D), William's R (R), Stochastic Oscillator K (K), Stochastic Oscillator D (D), Patterns (P).

Author	SMA	WMA	MOM	MACD	RSI	CCI	ROC	MAO	A/D	R	K	D	P	Other
Arévalo et al. (2017)													x	
Bustos et al. (2018)	x	x	x	x	x	x	x	x	x	x	x	x		
Dash and Dash (2016)	x		x		x		x			x	x	x		x
Dingli and Fournier (2017)			x											x
Ghanavati et al. (2016)	x	x	x		x	x			x	x	x	x		
Gunduz et al. (2017)	x	x		x	x		x			x	x	x		x
Guo et al. (2017)	x		x				x			x				x
Kim and Enke (2016)			x	x	x	x	x	x		x	x	x		x
Labiad et al. (2016)	x	x			x		x				x	x		x
Leito et al. (2016)													x	
Li, Tam, et al. (2016)														x
Liu et al. (2018)					x					x	x	x		
Mingyue et al. (2016)			x		x	x	x			x	x	x		x
Nelson et al. (2017)														x
Patel et al. (2015)	x	x	x	x	x	x			x	x	x	x		
Qiu and Song (2016)			x		x	x	x			x	x	x		x
Shynkevich et al. (2014)	x	x	x	x	x	x	x			x	x	x		x
Van Den Poel et al. (2016)	x	x		x		x				x	x	x		x
Wang, Xu, et al. (2018)	x	x	x	x	x	x	x							x
Weng et al. (2017)					x					x	x			
Yang et al. (2017)					x									x

Likewise, financial reports can help to estimate the value of the stock. The reports generated by the companies are composed of The Balance Sheet, The Income Statement, and The Cash Flow. In each of these, the company's result is reflected in the measured period. Those can help to estimate if the value of the company is increased or decreased. Moreover, just as in technical analysis, here are calculated indicators called financial ratios, which summarize the company's behavior numerically. Kamble (2018) used some financial ratios like Debt/equity Ratio, Price/Earnings Ratio, and Profit to earn Ratio to forecast the stock prices. Chai, Du, Lai, and Lee (2015) also used Price/Earnings Ratio as predictive input, in conjunction with New Loan/Market Capitalization Ratio.

The advantage of this information is that it is usually free of charge and is published by governments and companies. Moreover, since it tries to explain the reason for the movements, it makes its study interesting. However, their main disadvantage is the low frequency with which governments publish this data. Besides, price indexes and unemployment rates are usually published each month, which makes this information not so useful for predicting day-to-day price changes. In the case of companies, they publish their results on a quarterly, semi-annual, or annual basis, which makes a daily prediction even more difficult.

3.2. Unstructured inputs

The use of unstructured information is an additional challenge in predicting the stock market. This information must be preprocessed and converted to categorical or numerical information in order to be used as an input for the model. Textual unstructured inputs require text mining techniques, which extract news segments or opinions from social networks and can generate numerical representations that can predict stock prices. In addition, most of this information is not continuous but eventual, which changes the prediction approach of a time series, and becomes an if-then type analysis.

The analysis of the news is usually taken from three different sources: specialized media in finance, news in general, and news generated by the same company. Information provided in financial media is usually contextualized in the economic context, while news, in general, may not have a defined context, but directly impact the price of a stock. Several studies show that this kind of information is useful while predicting a stock price (Feuerriegel & Fehrer, 2016; Ghanavati, Wong, Chen, Wang, & Fong, 2016; Li, Chen, Jiang, Li, & Chen, 2016a; Liu, Zeng, Yang, & Carrio, 2018; Shynkevich, McGinnity, Coleman, & Belatreche, 2015). Feuerriegel and Fehrer (2016) worked with regulated ad hoc announcements in order to forecast the German stock market. They increased the accuracy benchmark of this problem by 5.66% using this kind of information in conjunction with the deep learning technique.

The study of social networks and search engines to predict the stock market is a relatively new field. In addition to the problems encountered in processing unstructured information such as the news, the volume of information to process is enormous, which can exceed millions of records, and cause computational challenges. As an example, Coyne, Madiraju, and Coelho (2018) processed 1,013,794 tweets and Pagolu, Reddy, Panda, and Majhi (2017) processed 2,500,000. This complexity implies that specialized techniques and powerful machines must be applied for processing. Also, unlike news, communications on social networks are not usually written in a standard format, such as titles, and may have spelling mistakes and emoticons (Porshnev, Redkin, & Karpov, 2015).

Many of the work done in social networks try to estimate whether the sentiment towards the company is positive or negative, and it has been proven that these metrics can be useful for predicting the stock market (Chakraborty, Pria, Rony, & Majumdar, 2018; Coyne et al., 2018; Dang & Duong, 2016; Li, Chan, Ou, & Ruifeng, 2017; Sun et al., 2017; Wang, Xu, & Zheng, 2018; Xu & Kešelj, 2014). Also, the analysis of the trends in search engines has proven useful (Hu, Tang, Zhang, & Wang, 2018; Weng, Ahmed, & Megahed, 2017).

4. Data pre-Processing

4.1. Natural language processing

Most of the works carried out for stock prediction use the Bag of Words technique to vectorize textual information and introduce it to a predictive algorithm (Khadjeh Nassirtoussi et al., 2014). This technique consist on representing the input document as a frequency vector, where the number of times a word appears is counted. Before constructing this frequency vector, the words called Stopwords are filtered. Those can be discarded because do not contribute significantly to the analysis, such as prepositions or articles. Shynkevich et al. (2015) processed news from specific subjects using this technique, converting each article in a 500 feature vector, and training a Support Vector machine with that data. They normalized the relevance of each word using TF-IDF (Term Frequency - Inverse Document Frequency), where not only the appearance of the words is counted, but an estimate is made of their relevance in the corpus.

Other relevant Natural Language Processing (NLP) technique is word2vec, proposed by Mikolov, Sutskever, Chen, Corrado, and Dean (2013). This is an Embeddings vectorization technique based on the use of a multilayer perceptron neuronal network. This kind of technique can adequately represent the distance or similarity between words. Liu et al. (2018) implemented a forecast model processing company related news with word2vec. They also integrated the information from the stock market, improving the prediction performance obtained just with TF-IDF.

4.2. Order reduction

Principal component analysis (PCA) is a well-known technique to summarize high-dimensional datasets and extract the most relevant features of the training inputs. Zhong and Enke (2017) forecasted daily direction of the S&P 500 Index ETF using as input 60 financial and economic features. They conclude that using PCA not only simplified the amount of data needed to train the models, but also increased the overall accuracy of the predictions. Singh and Srivastava (2017) also used PCA to speed up the model training without loosing accuracy on the forecast accuracy.

5. Modeling

Forecasting consists in making predictions of the future using present and past data. In the case of a time series $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$ the purpose is to predict future values $y_{t-1}, y_{t-2}, \ldots, y_{t-q}$ based on the data collected to date (Shumway & Stoffer, 2010).

Fig. 5 presents the proposed taxonomy to classify the range of models used in the analyzed studies. Given that we are dealing with the problem of prediction of the stock market as a problem of classification and not of regression, we propose to make the analysis in the following categories: Support Vector Machines (SVM), Artificial Neural Networks (ANN), Fuzzy Logic (FL), Genetic Algorithms (GA), Ensemble Models (E), Bayesian Models (B), Decision Trees (DT) and Deep Learning (DL). Deep learning models are a subset of neural networks, they were intentionally classified in this way to differentiate from traditional neural networks that do require feature engineering. Another Category (O) was defined

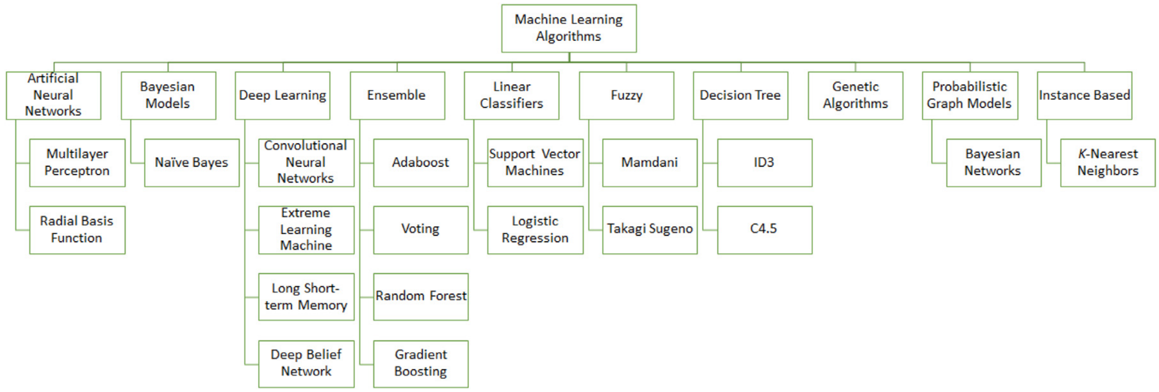

Fig. 5. Taxonomy of stock market prediction algorithms.

a subset of new algorithms that do not match former categories. Table 6 shows which modeling technique was used in each study.

5.1. Artificial neural networks

Artificial Neural Networks (ANN) are inspired by the functioning of the animal brain and are used as universal Turing machines. ANN is the most published Bio-Inspired technique and has been applied successfully in the stock market forecast. This family of techniques uses a large number of neural units, like the human brain, strongly interconnected among them. These connections may be either excitatory or inhibitory, affecting possible output values. Those values are added in a single value which is transformed in an output value using a threshold function.

Multilayer Perceptron (MLP) is the most used technique to forecast stocks. MLP is a feed-forward network, with an input layer, an output layer, and one or more hidden layers. Each layer adds non-linear learning capabilities. Usually, this network uses the Backpropagation technique for training, where the error of prediction is propagated from the output layer to the input layer, modifying interconnection weights. Coyne et al. (2018), Hu et al. (2018) Qiu and Song (2016), Mingyue, Cheng, and Yu (2016) and Zhong and Enke (2017) trained their models using the backpropagation technique for the MLP models.

Other common ANN technique is the radial basis function network (RBF). That kind of networks uses radial basis functions as activation functions. The RBF performs a linear combination of radial basis functions of the inputs in order to calculate the network output. Dash and Dash (2016) and Guo, Ye, Yang, and Zeng (2017) trained their models using the RBF technique.

5.1.1. Deep neural networks

Traditionally, machine learning algorithms require preprocessing and feature extraction before they can be trained (LeCun, Bengio, & Hinton, 2015). Sometimes the feature extraction requires the development of complex theory depending on the research field. Recently, given the advances in computational power, there are models like Deep Neural Networks that can use a cascade of multiple layers of non-linear processing units for automatic feature extraction and transformation. These layers allow the neural network to find non-linear relationships among data and improves the performance of learning from raw data (LeCun et al., 2015). The training algorithms have been improved to accept batches of data to speed up the learning. Moreover, there have been improvements in the technology itself, like GPU optimization for the training of these models (Coates et al., 2013), and new robust frameworks to build up these vast models (Abadi et al., 2016).

In the field of stock market prediction, few studies are using deep learning. Chong, Han, and Park (2017) used high-frequency intraday stock returns to forecast stock market using deep neural networks, and they compared them with other feature extraction components. They concluded that deep neural networks could extract additional information and therefore improve prediction performance.

Other related work reviewed was that of Singh and Srivastava (2017). In this article, they predicted stock prices using Google stock data and combining two powerful techniques: PCA and Deep Neural Network. They compared their results with a Radial Basis Function Neural Network and got an improvement in accuracy of 4.8%. In the study of Kraus and Feuerriegel (2017) the goal was to predict the stock prices movements based on financial disclosures. They used a corpus of 139 million words as a basis for the training of the algorithm of deep learning. They conclude that applying these novel techniques yields better performance than traditional techniques. In other study, Di Persio and Honchar (2016) compared Multi-Layer Perceptron (MLP), the Convolutional Neural Networks (CNN), and the Long Short-Term Memory (LSTM) for stock market movement forecast. They used S&P500 closing prices as the testbed and proposed a combination of wavelets and CNN for forecasting this index. They concluded that CNN outperforms classical neural networks for stock index forecasting.

As it can be noticed, deep learning techniques have become popular because they do not require performing the preprocessing of information beforehand. It is important to highlight, that there are only articles published between 2016 and 2018, which proves that this technique is novel and promising since these works have been compared with state of the art models and have shown better results. None of these works have used any additional input, as social networks or another type of economic indicator, then further publications of this topic are expected with the combination of multiple data sources.

5.2. Genetic algorithms

Genetic Algorithms (GA) mimic natural evolution processes to find an optimal solution to problems, like maximizing models accuracy. In these algorithms, solutions are combined, mutated an altered in each iteration, and the best solutions are selected for further iterations. It is used as a fitness function or objective function to evaluate which solutions are better than others. This algorithm starts with a random population, and it is expected that it finds an optimal solution.

Huang and Li (2017) used GA to predict the price movement of 10 stock in the Taiwan Stock Exchange. They performed a comparative study on the prediction performance, using linear regression and logistic regression models as a benchmark. Using temporal validation windows, they concluded that the proposed model

Table 6
Stock article by technique. Support Vector Machines (SVM), Deep Learning (DL), Artificial Neural Networks (ANN), Fuzzy Logic (FL), Genetic Algorithms (GA), Ensemble Models (E), Bayesian Models (B), Decision Trees (DT), Other Category (O) .

Author	SVM	DL	ANN	FL	GA	E	B	DT	O
Arévalo et al. (2017)									X
Bustos et al. (2018)	X		X						
Cen et al. (2017)			X				X		
Chai et al. (2015)	X								
Chakraborty et al. (2018)	X					X		X	
Coyne et al. (2018)			X						
Dang and Duong (2016)	X								
Dash and Dash (2016)			X						
Dingli and Fournier (2017)		X							
Di Persio and Honchar (2016)		X				X			
Feuerriegel and Fehrer (2016)		X							
Fischer and Krauss (2018)		X							
Ghanavati et al. (2016)				X					
Gonzalez et al. (2015)	X								
Hu et al. (2018)			X						
Huang and Li (2017)					X				
Huang et al. (2018)						X			
Kamble (2018)						X			
Kia et al. (2018)									X
Kim and Enke (2016)									X
Labiad et al. (2016)						X			
Leito et al. (2016)					X				
Li, Chen, et al. (2016a)						X			
Li, Tam, et al. (2016)	X								
Li et al. (2017)									X
Liu et al. (2018)		X							
Malagrino et al. (2018)							X		
McCluskey and Liu (2017)						X			
Mingyue et al. (2016)			X						
Nelson et al. (2017)		X							
Pagolu et al. (2017)	X								
Patel et al. (2015)						X			
Porshnev et al. (2015)	X								
Qiu and Song (2016)		X							
Ren et al. (2018)	X								
Shynkevich et al. (2014)	X								
Shynkevich et al. (2015)	X								
Sun et al. (2017)	X								
Tsantekidis et al. (2017)	X	X	X						
Van Den Poel et al. (2016)						X			
Wang, Liu, et al. (2018)	X								
Wang, Xu, et al. (2018)	X	X	X			X			
Weng et al. (2017)	X		X					X	
Xu and Kešelj (2014)	X								
Yang et al. (2017)						X			
Zhang et al. (2018)						X			
Zhong and Enke (2017)			X						
Zhou et al. (2018)								X	

GA-based system improved the accuracy of the prediction of the base models.

5.3. Fuzzy logic

Fuzzy Logic tries to mimic human reasoning. Fuzzy Logic allows the creation of soft if-then rules, where the premise use categories instead of exact values. Fuzzy Logic is a powerful tool when learning rules from human experts. Adaptive neuro-fuzzy inference system (ANFIS) is the algorithm most often used. This algorithm consists of a network of neuro-diffuse components from a training set to capture the interpretation of linguistic variables and gives as output a set of linear models. This technique has proven useful in environments such as systems control, but also shows the power to make predictions. Ghanavati et al. (2016) proposed a hybrid modeling technique, Fuzzy-Based Local Metric Learning (FuzyyML) extension based on SVM (FuzzyML-SVM), which combines Fuzzy Clustering with Support Vector Machines. They found that this hybrid technique outperforms SVM alone.

5.4. Ensembles

The best-known ensemble techniques are Boosting, Bagging, and Stacking. Boosting techniques correspond to meta-ensemble algorithms that train their models by changing the distribution of training data. This type of algorithms gives more weight to training examples that have not been well classified. The most significant difference with Bagging methods is that creating a new model in the ensemble requires that the training of the previous model has been carried out. A particular case is the technique of Boosted Trees, used by McCluskey and Liu (2017). This algorithm consists of training a set of decision trees, where each tree added to the ensemble specializes in classifying the examples in which the other trees had a high error. This specialization allows that as ensemble grows, the more performance it will be on the training set. Huang, Kong, Li, Yang, and Li (2018) modeled the stock market using the AdaBoost algorithm to improve a Naive Bayesian classifier into a robust classifier.

On the other hand, Bagging (or Bootstrap Aggregation) techniques do not change the data distribution. They create a ran-

Table 7
Performance comparison of the stock market forecast models.

Author	Output	Performance	Performance measurement
Arévalo et al. (2017)	Success Rate	-	
Bustos et al. (2018)	Up/Down Daily	0,72	Accuracy
Cen et al. (2017)	Buy, Hold, Sell	-	
Chai et al. (2015)	Up/Down Daily	0,79	Accuracy
Chakraborty et al. (2018)	Up/Down Daily	0,79	Accuracy,
Coyne et al. (2018)	Up/Down Daily	0,78	Accuracy
Dang and Duong (2016)	Up/Down 4 months	0,73	Accuracy
Dash and Dash (2016)	Up/Down Daily	0,8	Accuracy, F-Measure
Dingli and Fournier (2017)	Up/Down Next Month, Next Week	0,65	Accuracy
Di Persio and Honchar (2016)	Up/Down Daily	0,55	MSE, Accuracy
Feuerriegel and Fehrer (2016)	Up/Down on Disclosure	0,56	Accuracy
Fischer and Krauss (2018)	Up/Down Daily	0,54	Return, Standard Deviation, Sharpe Ratio, Accuracy
Ghanavati et al. (2016)	Up/Down Next Month, Next Week	0,85	Error Rate, F-measure
Gonzalez et al. (2015)	Up/Down Movement Weekly		Accuracy
Hu et al. (2018)	Up/Down Daily	0,89	Hit Ratio
Huang and Li (2017)	Up/Down Intraday	55-100%	Accuracy
Huang et al. (2018)	Buy, Hold, Sell	1-51%	profit ratio
Kamble (2018)	Buy, Hold, Sell	66.8%	Accuracy
Kia et al. (2018)	Up/Down Daily	56.78	Accuracy
Kim and Enke (2016)	Success Rate	66%	Success Rate
Labiad et al. (2016)	Up/Down Intraday	0,95	Accuracy
Leito et al. (2016)	Success Rate	72,8%	Success Rate
Li, Chen, et al. (2016)	Up/Down Intraday 20 minute window	62-65%	Accuracy
Li, Tam, et al. (2016)	Up/Down Daily	53.5% and 66.6%	Accuracy
Li et al. (2017)	Up/Down Daily	52-76%	Accuracy
Liu et al. (2018)	Up/Down Daily	55.44%	Accuracy, F1-Score
Malagrino et al. (2018)	Up/Down Daily	71-78%	Accuracy
McCluskey and Liu (2017)	Up/Down Daily	0,59	Accuracy,
Ming, Wong, Liu, and Chiang (2015)	Up/Down Daily	55.7%	Accuracy
Mingyue et al. (2016)	Up/Down Daily	86.39%	Hit Ratio
Nelson et al. (2017)	Up/Down Daily	55.9%	Accuracy
Pagolu et al. (2017)	Up/Down Daily	71.82%	Accuracy
Patel et al. (2015)	Up/Down Daily	83.56%.	accuracy
Porshnev et al. (2015)	Up/Down Daily	59.69%	accuracy
Qiu and Song (2016)	Up/Down Daily	81.27%.	Hit Ratio
Ren et al. (2018)	Up/Down Daily	0,89	accuracy
Shynkevich et al. (2014)	Up/Down Daily	91.9%	accuracy
Shynkevich et al. (2015)	Up/Down Daily	79.59%	accuracy
Sun et al. (2017)	Up/Down,Flat Daily	57.3%	accuracy
Tsantekidis et al. (2017)	Up/Down, intraday	75.92	Precision*,
Wang, Liu, et al. (2018)	Up/Down Daily	58.30%	accuracy
Wang, Xu, et al. (2018)	Up/Down Daily	79.4%	F1 Precision Recall Accuracy AUC
Weng et al. (2017)	5 different One-day-ahead	0,85	Accuracy,
Xu and Kešelj (2014)	Up/Down Daily	58.9%	Accuracy
Yang et al. (2017)	Up/Down Daily	0,85	Accuracy
Zhong and Enke (2017)	Up/Down Daily	58.1%	Accuracy, Return
Zhou et al. (2018)	Up/Down Daily, Weekly	66.67%	Accuracy

dom subset of the training data. The most relevant Bagging algorithm is the Random Forest (RF). This algorithm consists of training a determined number of decision trees, which use a subset of randomly selected input variables. By far, this is the most popular ensemble reported in the literature for predicting the stock market (Kamble, 2018; Labiad, Berrado, & Benabbou, 2016; Patel, Shah, Thakkar, & Kotecha, 2015; Van Den Poel, Chesterman, Koppen, & Ballings, 2016; Zhang, Cui, Xu, Li, & Li, 2018). Patel et al. (2015) concluded that the Random Forest technique improves the predictions made by the SVMs.

Finally, stacking corresponds to techniques that allow combining multiple models already trained. Unlike the described techniques of Boosting and Bagging, which used mainly decision trees, Stacking allows combining models of different types. This stacking makes it flexible when mixing predictors, even those that can use different sources of information for prediction. In the article by Yang et al. (2017), they combined ensemble of SVM, RF, and Logit AdaBoost using average voting to improve the performance of the classification models.

5.5. Support vector machines

Support Vector Machines are used for classification and regression analysis. This algorithm guarantees to find a hyperplane that maximizes the distance between two sets of data. Using the kernel trick, the Support Vector Machines can find hyperplanes in higher-level dimensions. The SVM technique is the most common among linear separation algorithms since it is virtually parameter free and has shown that it can have the same or better performance than other more complex algorithms.

Table 6 shows that the most popular method is SVM. This method has few hyperparameters to tune, while it leads to high levels of forecast accuracy. Biologically inspired methods, such as ANN and fuzzy logic, are popular too. Those can deal with non-linear relationships, which makes them powerful methods for forecasting. However, in general, these methods create black box models, which makes them difficult to interpret. In particular, deep learning techniques have gained popularity recently, due to advances in the ease of computation and the refinements that have been made on their execution.

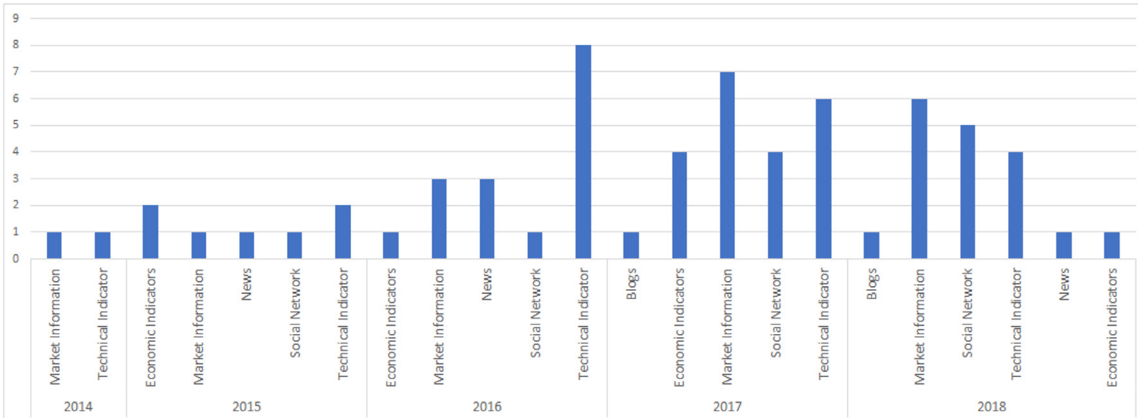

Fig. 6. Input type reported in the last years.

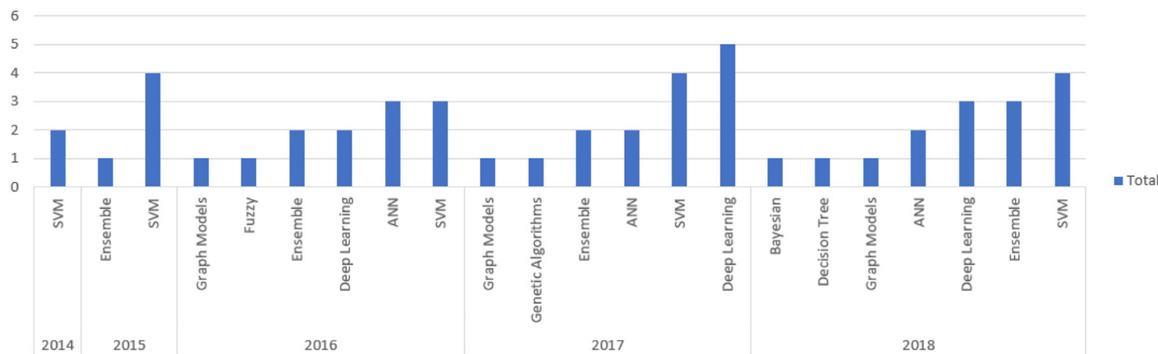

Fig. 7. Algorithms reported in the last years.

5.6. Rough sets

Rough set is a mathematical technique to approximate uncertainty data in the for of lower and upper limits. Kim and Enke (2016) forecasted the Korea Composite Stock Price Index 20 0 (KOSPI 20 0) using a mixture of techniques, like genetic algorithms and trading strategies, based on Rough Sets, and compared the profitability of benchmark trading strategies like Buy-and-hold, where Rough Sets outperformed the other strategies.

6. Evaluation

The problem of stock market prediction can also be classified by the type of output to be estimated. Two major problems can be addressed, classification, and regression. For the former, it is usually simplified to return categories defined as up / down. For the second, the output is the numerical prediction of how much a stock is going to go up or down. This article surveys only the articles focused in predicting stock market movement direction, addressed as a classification problem. Table 7 shows the output type reported in the article, its type of output, and the performance measurement described. Some studies without performance data were omitted.

On the classification problem, the performance metrics are usually accuracy metrics. Accuracy is a percentage measure of how many hits the algorithm has concerning the complete test set. However, it is not the only one, since it does not consider the importance of false positives (type 1 error) or false negatives (type 2 error). Precision is the percentage measure of the true positive over the total predicted positives, and it is an important indicator when the cost of a false positive is high (Gunduz, Yaslan, & Cataltepe, 2017; Tsantekidis et al., 2017; Wang, Xu, et al., 2018). Recall calculates the percentage of true positives over the number of total positives, and it is useful when the cost of a false negative is high (Wang, Xu, et al., 2018). F1 score is a measure that mixes precision and recall, balancing the importance of false positive and false negative (Dash & Dash, 2016; Ghanavati et al., 2016; Liu et al., 2018; Wang, Xu, et al., 2018).

In addition, some articles report the performance as the return percentage or profit ratio, where a trading technique using the prediction algorithm is back-tested, to predict how profitable are the predictions (Fischer & Krauss, 2018; Huang et al., 2018; Zhong & Enke, 2017).

7. Overall analysis

Fig. 6 shows that the technical indicators and raw market information are the most popular type of input to predict the stock market. Since the technical indicators have been refined empirically for years by stock market analysts, they capture the essence of the trends and patterns of the time series of the stock market. However, an interest in the analysis of social networks for the prediction of the stock market has been aroused, and it is expected that in the coming years, more articles that mix the technical analysis and the analysis of social networks will be published.

Fig. 7 shows the usage of each machine learning technique over the last years. SVM remains the most popular technique to date. However, deep learning techniques have drawn attention to the scientific community recently.

Fig. 8 presents the performance results of the models and compares them to the country studied. Frontier and emerging markets models show better accuracy than the developed countries. It is possible that the randomness and complexity of developed countries limit the possibility of creating accurate models. Multi-

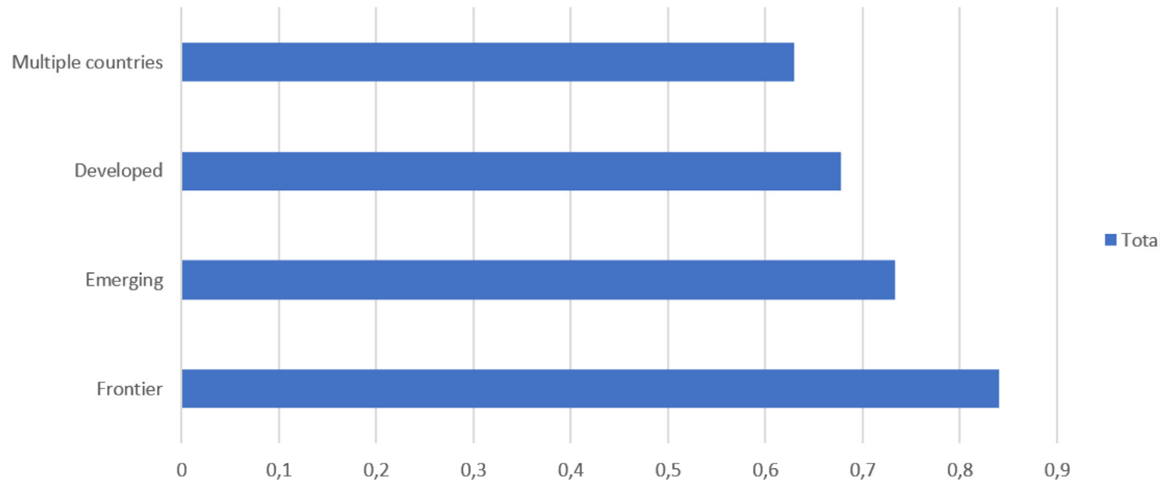

Fig. 8. Accuracy by country type.

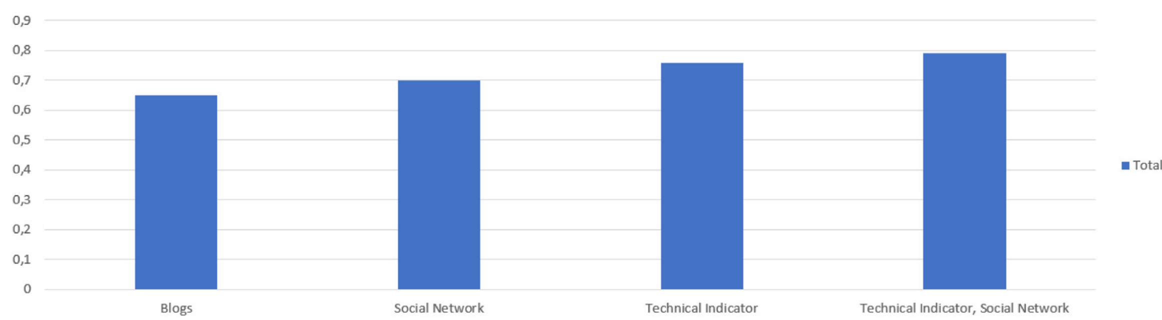

Fig. 9. Accuracy by input type.

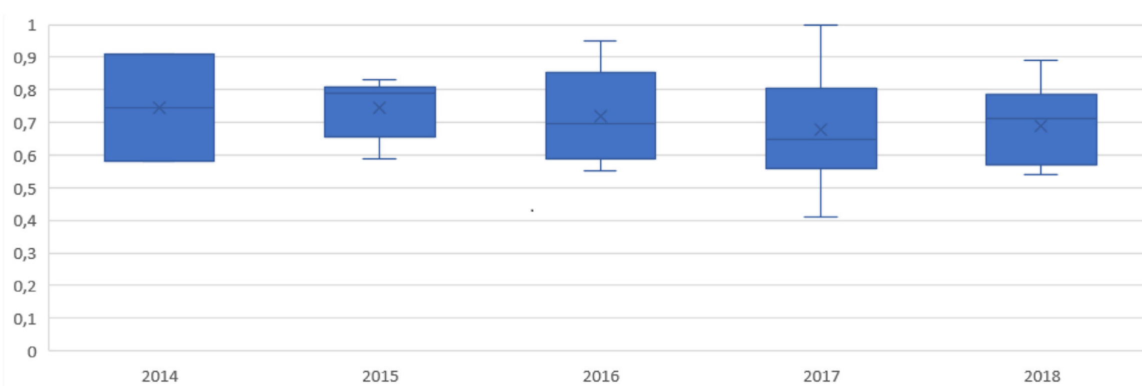

Fig. 10. Accuracy by publication year.

ple countries indicate that the study applied the model in more than one country.

Furthermore, Fig. 9 shows the cross-comparative between the performance results of the models and the type of input used by the model. It can be concluded that as more information is combined, the models perform better. The data from blogs and social networks alone do not perform better (on average) than just using technical indicators. However, if they are combined with them, the performance of the prediction can be improved.

Finally, Fig. 10 makes a new finding of the evolution of the models over time. In general, the works reported in the literature do not show an improvement in the performance of the models as time passes. It is found that, on average, the published works have a precision between 0.7 and 0.8 in the last five years. However, it is also found that in 2017 there have been reports of very uneven performance. This evolution on the accuracy may be due to the difference in performance in markets not previously analyzed.

8. Conclusion

This article shows an updated review of the literature on stock market prediction. It is focused on the works between 2014 and 2018 that perform prediction work of the stock market as a classification problem. For this, 52 studies were characterized and described in detail depending on the type of inputs and outputs they use, as well as the type of modeling algorithm they apply.

The most popular source of information to forecast the stock market are technical indicators. The technical indicators have proven to be the most predictive data of all. However, as it was analyzed in Section 7, the information of the social networks improves the performance of the models. So far, most works on social networks have focused on the analysis of sentiment. These works can be enriched by analyzing the topics discussed in social networks, in such a way that the models can have a complete and more precise idea of what is happening in the country and the world.

There has been an increase in popularity of sophisticated machine learning algorithms, such as ensemble models and deep learning. The ensemble models have shown high predictive power, even in some comparative works they have performed better than other techniques such as SVM and ANN. Deep learning models, in general, have not outperformed traditional models. It is possible that the data sets with which these algorithms have been trained is not sufficient to generate an adequate prediction.

Then it is to be expected that future work can be focused on finding new sources of information that complement the technical analysis to predict stock markets. For example, in addition to analyzing sentiment about networks, it can be complemented with the analysis of the topics that are spoken on the networks, which leads to the models having a more complex idea of what is happening in the world. Additionally, given the success in other types of problems of the free models of feature engineering, in the future, more articles are expected to find optimal technical indicators automatically.

Finally, there are missing comparative works among the forecasting on developed, emerging, and frontier markets. It is crucial to find what kind of information and which modeling techniques are appropriate for each of the stock markets.

It is important to mention that there are some limitations to this review. First, it may have missed relevant articles that were not indexed in the selected databases. Second, though the definition of the keywords involved much work and several iterations, we could have missed some studies that used less common language terms to refer stock market prediction. Given the nature of this review, viz. an overview of academic literature, we are aware that much work done in the development of industrial software products is not reflected.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

O Bustos: Conceptualization, Methodology, Software. **A. Pomares-Quimbaya:** Data curation, Writing - original draft.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning.. In *OSDI: Vol. 16* (pp. 265–283).

Appel, G. (1985). *The moving average convergence-divergence trading method: Advanced version*. Scientific Investment Systems.

Arévalo, R., García, J., Guijarro, F., & Peris, A. (2017). A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting. *Expert Systems with Applications, 81*, 177–192. doi:10.1016/j.eswa.2017.03.028. http://www.sciencedirect.com/science/article/pii/S0957417417301823.

Atsalakis, G., & Valavanis, K. (2009). Surveying stock market forecasting techniques - part II: Soft computing methods. *Expert Systems with Applications, 36*(3 PART 2), 5932–5941. doi:10.1016/j.eswa.2008.07.006.

Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications, 19*(8), 1165–1195. doi:10.1007/s00521-010-0362-z.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8.

Boyacioglu, M. A., & Avci, D. (2010). An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the istanbul stock exchange. *Expert Systems with Applications, 37*(12), 7908–7912. doi:10.1016/j.eswa.2010.04.045. http://www.sciencedirect.com/science/article/pii/S0957417410003453.

Bustos, O., Pomares, A., & Gonzalez, E. (2018). A comparison between SVM and multilayer perceptron in predicting an emerging financial market: Colombian stock market. In *2017 Congreso internacional de innovacion y tendencias en ingenieria, CONIITI 2017 - conference proceedings: vol. 2018-January* (pp. 1–6). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/CONIITI.2017.8273335.

Cavalcante, R., Brasileiro, R., Souza, V., Nobrega, J., & Oliveira, A. (2016). Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications, 55*, 194–211. doi:10.1016/j.eswa.2016.02.006.

Cen, L., Ruta, D., & Ruta, A. (2017). Using recommendations for trade returns prediction with machine learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10352 LNAI*, 718–727. doi:10.1007/978-3-319-60438-1_70.

Chai, J., Du, J., Lai, K., & Lee, Y. (2015). A hybrid least square support vector machine model with parameters optimization for stock forecasting. *Mathematical Problems in Engineering, 2015*. doi:10.1155/2015/231394.

Chakraborty, P., Pria, U., Rony, M., & Majumdar, M. (2018). Predicting stock movement using sentiment analysis of Twitter feed. In *2017 6th International conference on informatics, electronics and vision and 2017 7th international symposium in computational medical and health technology, ICIEV-ISCMHT 2017: vol. 2018-January* (pp. 1–6). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/ICIEV.2017.8338584.

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.

Coates, A., Huval, B., Wang, T., Wu, D., Catanzaro, B., & Andrew, N. (2013). Deep learning with cots HPC systems. In *International conference on machine learning* (pp. 1337–1345).

Coyne, S., Madiraju, P., & Coelho, J. (2018). Forecasting stock prices using social media analysis. In *Proceedings - 2017 IEEE 15th international conference on dependable, autonomic and secure computing, 2017 IEEE 15th international conference on pervasive intelligence and computing, 2017 IEEE3rd international conference on big data intelligence and computing and 2017 IEEE cyber science and technology congress, dasc-picom-datacom-cyberscitec 2017: 2018–January* (pp. 1031–1038). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/DASC-PICom-DataCom-CyberSciTec.2017.169.

Dang, M., & Duong, D. (2016). Improvement methods for stock market prediction using financial news articles. In *NICS 2016 - Proceedings of 2016 3rd national foundation for science and technology development conference on information and computer science* (pp. 125–129). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/NICS.2016.7725636.

Dash, R., & Dash, P. (2016). A comparative study of radial basis function network with different basis functions for stock trend prediction. In *2015 IEEE power, communication and information technology conference, PCITC 2015 - Proceedings* (pp. 430–435). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/PCITC.2015.7438204.

Di Persio, L., & Honchar, O. (2016). Artificial neural networks architectures for stock price prediction: Comparisons and applications. *International Journal of Circuits, Systems and Signal Processing, 10*, 403–413.

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction.. In *IJCAI* (pp. 2327–2333).

Dingli, A., & Fournier, K. (2017). Financial time series forecasting - A deep learning approach. *International Journal of Machine Learning and Computing, 7*(5), 118–122. doi:10.18178/ijmlc.2017.7.5.632.

Fama, E. F. (1995). Random walks in stock market prices. *Financial Analysts Journal, 51*(1), 75–80.

Feuerriegel, S., & Fehrer, R. (2016). Improving decision analytics with deep learning: The case of financial disclosures. *24th European conference on information systems, ECIS 2016*. Association for Information Systems.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270*(2), 654–669. doi:10.1016/j.ejor.2017.11.054.

George Atsalakis, K. V. (2009). Surveying stock market forecasting techniques - Part II: Soft computing methods. *Expert Systems with Applications*.

Ghanavati, M., Wong, R., Chen, F., Wang, Y., & Fong, S. (2016). A generic service framework for stock market prediction. In *Proceedings - 2016 IEEE international conference on services computing, SCC 2016* (pp. 283–290). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/SCC.2016.44.

Gonzalez, R., Padilha, C., & Barone, D. (2015). Ensemble system based on genetic algorithm for stock market forecasting. In *2015 ieee congress on evolutionary computation, CEC 2015 - Proceedings* (pp. 3102–3108). Institute of Electrical and Electronics Engineers Inc.. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84963553395. doi:10.1109/CEC.2015.7257276. Cited By 4; Conference of IEEE Congress on Evolutionary Computation, CEC 2015 ; Conference Date: 25 May 2015 Through 28 May 2015; Conference Code: 118157.

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.

Guo, Z., Ye, W., Yang, J., & Zeng, Y. (2017). Financial index time series prediction based on bidirectional two dimensional locality preserving projection. In *2017 IEEE 2nd international conference on big data analysis, ICBDA 2017* (pp. 934–938). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/ICBDA.2017.8078775.

Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with Google trends. *Neurocomputing*. doi:10.1016/j.neucom.2018.01.038.

Huang, C.-F., & Li, H.-C. (2017). An evolutionary method for financial forecasting in microscopic high-speed trading environment. *Computational Intelligence and Neuroscience, 2017*. doi:10.1155/2017/9580815.

Huang, Q., Kong, Z., Li, Y., Yang, J., & Li, X. (2018). Discovery of trading points based on Bayesian modeling of trading rules. *World Wide Web*, 1–18. doi:10.1007/s11280-018-0534-9.

Ince, H., & Trafalis, T. B. (2017). A hybrid forecasting model for stock market prediction.. *Economic Computation & Economic Cybernetics Studies & Research, 51*(3).

Kamble, R. (2018). Short and long term stock trend prediction using decision tree. In *Proceedings of the 2017 international conference on intelligent computing and control systems, ICICCS 2017: vol. 2018-January* (pp. 1371–1375). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/ICCONS.2017.8250694.

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. doi:10.1016/j.asoc.2012.09.024. http://www.sciencedirect.com/science/article/pii/S1568494612004449.

Khadjeh Nassirtoussi, A., Aghabozorgi, S., Ying Wah, T., & Ngo, D. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications, 41*(16), 7653–7670. doi:10.1016/j.eswa.2014.06.009.

Kia, A., Haratizadeh, S., & Shouraki, S. (2018). A hybrid supervised semi-supervised graph-based model to predict one-day ahead movement of global stock markets and commodity prices. *Expert Systems with Applications, 105*, 159–173. doi:10.1016/j.eswa.2018.03.037.

Kim, Y., & Enke, D. (2016). Developing a rule change trading system for the futures market using rough set analysis. *Expert Systems with Applications, 59*, 165–173. doi:10.1016/j.eswa.2016.04.031. http://www.sciencedirect.com/science/article/pii/S0957417416302044.

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*.

Krollner, B., Vanstone, B., & Finnie, G. (2010). Financial time series forecasting with machine learning techniques: A survey. In *Proceedings of the 18th European symposium on artificial neural networks - Computational intelligence and machine learning, ESANN 2010, Bruges* (pp. 25–30).

Labiad, B., Berrado, A., & Benabbou, L. (2016). Machine learning techniques for short term stock movements classification for moroccan stock exchange. *Sita 2016 - 11th International conference on intelligent systems: Theories and applications*. Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/SITA.2016.7772259.

Lambert, D. R. (1983). Commodity channel index: Tool for trading cyclic trends. *Technical Analysis of Stocks & Commodities, 1*, 47.

Lane, G. C. (1985). Lanes stochastics: The ultimate oscillator. *Journal of Technical Analysis, 21*, 37–42.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Leito, J., Neves, R. F., & Horta, N. (2016). Combining rules between pips and sax to identify patterns in financial markets. *Expert Systems with Applications, 65*, 242–254. doi:10.1016/j.eswa.2016.08.032. http://www.sciencedirect.com/science/article/pii/S0957417416304249.

Li, B., Chan, K., Ou, C., & Ruifeng, S. (2017). Discovering public sentiment in social media for predicting stock movement of publicly listed companies. *Information Systems, 69*, 81–92. doi:10.1016/j.is.2016.10.001.

Li, Q., Chen, Y., Jiang, L., Li, P., & Chen, H. (2016). A tensor-based information framework for predicting the stock market. *ACM Transactions on Information Systems, 34*(2). doi:10.1145/2838731.

Li, Y., & Ma, W. (2010). Applications of artificial neural networks in financial economics: A survey. In *Proceedings - 2010 international symposium on computational intelligence and design, ISCID 2010, Hangzhou: Vol. 1* (pp. 211–214).

Li, Z., Tam, V., & Yeung, L. (2016). Combining cloud computing, machine learning and heuristic optimization for investment opportunities forecasting. In *2016 IEEE congress on evolutionary computation, CEC 2016* (pp. 3469–3476). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/CEC.2016.7744229.

Liu, Y., Zeng, Q., Yang, H., & Carrio, A. (2018). Stock price movement prediction from financial news with deep learning and knowledge graph embedding. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11016 LNAI*, 102–113. doi:10.1007/978-3-319-97289-3_8.

Malagrino, L., Roman, N., & Monteiro, A. (2018). Forecasting stock market index daily direction: A Bayesian network approach. *Expert Systems with Applications, 105*, 11–22. doi:10.1016/j.eswa.2018.03.039.

McCluskey, J., & Liu, J. (2017). US financial market forecasting using data classification with features from global markets. In *2017 2nd International conference on image, vision and computing, ICIVC 2017* (pp. 965–969). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/ICIVC.2017.7984698.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Ming, F., Wong, F., Liu, Z., & Chiang, M. (2015). Stock market prediction from WSJ: Text mining via sparse matrix factorization. In *Proceedings - IEEE international conference on data mining, ICDM: 2015–January* (pp. 430–439). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/ICDM.2014.116.

Mingyue, Q., Cheng, L., & Yu, S. (2016). Application of the artifical neural network in predicting the direction of stock market index. In *Proceedings - 2016 10th international conference on complex, intelligent, and software intensive systems, CISIS 2016* (pp. 219–223). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/CISIS.2016.115.

Nardo, M., Petracco-Giudici, M., & Naltsidis, M. (2016). Walking down wall street with a tablet: A survey of stock market predictions using the web. *Journal of Economic Surveys, 30*(2), 356–369. doi:10.1111/joes.12102.

Nazário, R. T. F., e Silva, J. L., Sobreiro, V. A., & Kimura, H. (2017). A literature review of technical analysis on stock markets. *The Quarterly Review of Economics and Finance, 66*, 115–126. doi:10.1016/j.qref.2017.01.014. http://www.sciencedirect.com/science/article/pii/S1062976917300443.

Nelson, D., Pereira, A., & De Oliveira, R. (2017). Stock market's price movement prediction with LSTM neural networks. In *Proceedings of the international joint conference on neural networks: vol. 2017-May* (pp. 1419–1426). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/IJCNN.2017.7966019.

Pagolu, V., Reddy, K., Panda, G., & Majhi, B. (2017). Sentiment analysis of Twitter data for predicting stock market movements. In *International conference on signal processing, communication, power and embedded system, scopes 2016 - Proceedings* (pp. 1345–1350). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/SCOPES.2016.7955659.

Park, C.-H., & Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys, 21*(4), 786–826.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications, 42*(1), 259–268. doi:10.1016/j.eswa.2014.07.040.

Porshnev, A., Redkin, I., & Karpov, N. (2015). Modelling movement of stock market indexes with data from emoticons of Twitter users. *Communications in Computer and Information Science, 505*, 297–306. doi:10.1007/978-3-319-25485-2_10.

Preethi, G., & Santhi, B. (2012). Stock market forecasting techniques: A survey. *Journal of Theoretical and Applied Information Technology, 46*(1), 24–30.

Qiu, M., & Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PLoS ONE, 11*(5). doi:10.1371/journal.pone.0155133.

Ren, R., Wu, D., & Liu, T. (2018). Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Systems Journal*. doi:10.1109/JSYST.2018.2794462.

Ripon, K., & Rajon, S. (2016). Stock market forecast using bio-inspired computing. In *International conference on control, decision and information technologies, CODIT 2016* (pp. 523–529). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/CoDIT.2016.7593617.

Rubén, Aguilar-Rivera, & Manuel Valenzuela-Rendón, J. R.-O. (2015). Genetic algorithms and darwinian approaches in financial applications: A survey. *Expert Systems with Applications*.

Sapankevych, N., & Sankar, R. (2009). Time series prediction using support vector machines: A survey. *IEEE Computational Intelligence Magazine, 4*(2), 24–38. doi:10.1109/MCI.2009.932254.

Shumway, R. H., & Stoffer, D. S. (2010). *Time series analysis and its applications: With R examples*. Springer Science & Business Media.

Shynkevich, Y., McGinnity, T., Coleman, S., & Belatreche, A. (2015). Stock price prediction based on stock-specific and sub-industry-specific news articles. In *Proceedings of the international joint conference on neural networks: vol. 2015-September*. Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/IJCNN.2015.7280517.

Shynkevich, Y., McGinnity, T., Coleman, S., Li, Y., & Belatreche, A. (2014). Forecasting stock price directional movements using technical indicators: Investigating window size effects on one-step-ahead forecasting. In *IEEE/IAFE Conference on computational intelligence for financial engineering, proceedings (CIFER)* (pp. 341–348). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/CIFEr.2014.6924093.

Singh, R., & Srivastava, S. (2017). Stock prediction using deep learning. *Multimedia Tools and Applications, 76*(18), 18569–18584.

Sun, T., Wang, J., Zhang, P., Cao, Y., Liu, B., & Wang, D. (2017). Predicting stock price returns using microblog sentiment for chinese stock market. In *Proceedings - 2017 3rd international conference on big data computing and communications, bigcom 2017* (pp. 87–96). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/BIGCOM.2017.59.

Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing Journal, 38*, 788–804. doi:10.1016/j.asoc.2015.09.040.

Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., & Iosifidis, A. (2017). Using deep learning to detect price change indications in financial markets. In *25th European signal processing conference, EUSIPCO 2017: vol. 2017-January* (pp. 2511–2515). Institute of Electrical and Electronics Engineers Inc.. doi:10.23919/EUSIPCO.2017.8081663.

Tüfekci, P. (2016). Predicting the direction of movement for stock price index using machine learning methods. *Advances in Intelligent Systems and Computing, 427*, 477–492. doi:10.1007/978-3-319-29504-6_45.

Van Den Poel, D., Chesterman, C., Koppen, M., & Ballings, M. (2016). Equity price direction prediction for day trading: Ensemble classification using technical analysis indicators with interaction effects. In *2016 IEEE congress on evolutionary computation, cec 2016* (pp. 3455–3462). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/CEC.2016.7744227.

Vui, C., Soon, G., On, C., Alfred, R., & Anthony, P. (2013). A review of stock market prediction with artificial neural network (ann). In *Proceedings - 2013 ieee international conference on control system, computing and engineering, iccsce 2013* (pp. 477–482). https://www.scopus.com/inward/record.uri?eid=2-s2.0-84894141707. doi:10.1109/ICCSCE.2013.6720012. Penang.

Wang, J., Liu, Z., Shang, W., & Wang, S. (2018). An enhanced LGSA-SVM for s&p 500 index forecast. In *Proceedings - 2017 IEEE international conference on big data, big data 2017: 2018–January* (pp. 4176–4183). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/BigData.2017.8258441.

Wang, Q., Xu, W., & Zheng, H. (2018). Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles. *Neurocomputing, 299*, 51–61. doi:10.1016/j.neucom.2018.02.095.

Wen, Q., Yang, Z., Song, Y., & Jia, P. (2010). Automatic stock decision support system based on box theory and {SVM} algorithm. *Expert Systems with Applications, 37*(2), 1015–1022. doi:10.1016/j.eswa.2009.05.093. http://www.sciencedirect.com/science/article/pii/S0957417409005107.

Weng, B., Ahmed, M., & Megahed, F. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems with Applications, 79*, 153–163. doi:10.1016/j.eswa.2017.02.041.

Wilder Jr, J. W. (1986). The relative strength index. *Journal of Technical Analysis of Stocks and Commodities, 4*, 343–346.

Williams, L. (1985). The ultimate oscillator. *Technical Analysis of Stocks and Commodities, 3*(4), 140–141.

Xing, F., Cambria, E., & Welsch, R. (2018). Natural language based financial forecasting: A survey. *Artificial Intelligence Review, 50*(1), 49–73. doi:10.1007/s10462-017-9588-9.

Xu, F., & Kešelj, V. (2014). Collective sentiment mining of microblogs in 24-hour stock price movement prediction. In *Proceedings - 16th IEEE conference on business informatics, CBI 2014: vol. 2* (pp. 60–67). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/CBI.2014.37.

Yang, J., Rao, R., Hong, P., & Ding, P. (2017). Ensemble model for stock price movement trend prediction on different investing periods. In *Proceedings - 12th international conference on computational intelligence and security, CIS 2016* (pp. 358–361). Institute of Electrical and Electronics Engineers Inc.. doi:10.1109/CIS.2016.86.

Yoo, P., Kim, M., & Jan, T. (2005). Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In *Proceedings - international conference on computational intelligence for modelling, control and automation, CIMCA 2005 and international conference on intelligent agents, web technologies and internet, Vienna: vol. 2* (pp. 835–841).

Zhang, J., Cui, S., Xu, Y., Li, Q., & Li, T. (2018). A novel data-driven stock price trend prediction system. *Expert Systems with Applications, 97*, 60–69. doi:10.1016/j.eswa.2017.12.026.

Zhong, X., & Enke, D. (2017). A comprehensive cluster and classification mining procedure for daily stock market return forecasting. *Neurocomputing, 267*, 152–168. doi:10.1016/j.neucom.2017.06.010.

Zhou, P.-Y., Chan, K., & Ou, C. (2018). Corporate communication network and stock price movements: Insights from data mining. *IEEE Transactions on Computational Social Systems, 5*(2), 391–402. doi:10.1109/TCSS.2018.2812703.