IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 11, 2021, accepted January 18, 2021, date of publication February 9, 2021, date of current version February 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3058133

Artificial Intelligence Applied to Stock Market Trading: A Review

FERNANDO G. D. C. FERREIRA[1], AMIR H. GANDOMI[2], (Senior Member, IEEE), AND RODRIGO T. N. CARDOSO[1]

[1]Department of Mathematical and Computational Modeling, Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte 30510-000, Brazil
[2]Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia

Corresponding author: Amir H. Gandomi (gandomi@uts.edu.au)

This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) under Grant 001 (Proc. 88881.361790/2019-01).

ABSTRACT The application of Artificial Intelligence (AI) to financial investment is a research area that has attracted extensive research attention since the 1990s, when there was an accelerated technological development and popularization of the personal computer. Since then, countless approaches have been proposed to deal with the problem of price prediction in the stock market. This paper presents a systematic review of the literature on Artificial Intelligence applied to investments in the stock market based on a sample of 2326 papers from the Scopus website between 1995 and 2019. These papers were divided into four categories: portfolio optimization, stock market prediction using AI, financial sentiment analysis, and combinations involving two or more approaches. For each category, the initial introductory research to its state-of-the-art applications are described. In addition, an overview of the review leads to the conclusion that this research area is gaining continuous attention and the literature is becoming increasingly specific and thorough.

INDEX TERMS Computational finance, algotradings, artificial intelligence, finance.

I. INTRODUCTION

Beginning in the 1990s with introduction of computational methods in finance, much research has focused on applying Artificial Intelligence (AI) to financial investments in the stock market. The main advantages of using computational approaches to automate the financial investment process include the elimination of "momentary irrationality" or decisions made based on emotions, ability to recognize and explore patterns that are looked over by humans, and immediate consumption of information in real-time. This area of knowledge has become known as Computational Finance.

More recently, within computational finance, there is increasing use of and research on AI techniques applied in financial investments. Although a computer conducts the vast majority of hedge fund trades in an automated way, 90% of these operations are still performed by a hardcoded procedure [12]. Thus, the ever-increasing application of artificial intelligence still has great potential for development.

Generally, AI is applied to finance in three different areas: the optimization of financial portfolios, prediction of future prices or trends in financial assets, and sentiment analysis of news or social media comments about the assets or companies. Despite the differences and peculiarities of each area, some works have proposed combinations of techniques from the different areas. Some other studies in the area of computational finance include the control of dynamic systems applied to the financial market [8], investor behavioral analysis [77], network analysis [37], [82] and clustering of financial assets [1]. Reference [98] relates the calibrated volatility of options to the movements in futures prices in the Taiwan stock market. It calculates a correlation of approximately -0.9 and concludes that the volatility of options can be used for the prediction of futures prices.

The presented work analyzes the development of each of these areas from its initiation to its state-of-the-art advancements. For this, a sample of works on artificial intelligence published between 1995-2019 was collected for detailed analysis and comparison.

This paper is structured as follows: Section II presents some general information extracted from the works on the application of AI to finance. Section III discusses the development of the portfolio optimization area and state-of-the-art applications. Section IV describes forecasting of

The associate editor coordinating the review of this manuscript and approving it for publication was Hiram Ponce.

assets prices and trends. Section V overviews the works in the area of sentiment analysis applied to news or comments on social media about these assets, while Section VI discusses the works that combine at least two of the three major areas mentioned. Finally, Section VII summarizes and presents final considerations about the development and the state-of-the-art advancements in all these areas.

II. OVERVIEW

The analysis of the works that deal with the application of artificial intelligence to financial investments is based on a sample of works taken from Scopus. The query, performed on May 27, 2020, returned a total of 2326 documents, including journal and conference papers published between 1995 and 2019.

Table 1 presents the most cited papers on Artificial Intelligence applied to financial investments. The oldest paper is from 2000 [13], and the most recent is from 2015 [63]. Therefore, it can be observed that this is a relatively new research area that is gaining increasing attention. Among the papers showed in Table 1, there are no more recent documents than 2015 as expected since they had less time to be studied by other researchers, and newer papers tend to be more specific.

TABLE 1. Most cited documents.

Article	Cited by
[9]	2069
[13]	464
[36]	462
[74]	352
[42]	278
[20]	278
[63]	237
[22]	235
[80]	221
[29]	217
[6]	213
[47]	204
[18]	197
[87]	196
[24]	188

The most cited papers in each year, from 1995 to 2019, are shown in Table 2, revealing that papers up to 1998 were not cited much considering that the area acquired more interest with the rapid improvement and popularization of computers in the following years. Again, a low amount of citations for 2019 was expected.

Table 3 shows the conferences or journals in which the AI applied in investment are most published and it reveals that most of the papers were published in *Lecture Notes In Computer Science*, which presents a much larger number of documents than the other sources, with 180 articles, followed by *Expert Systems with Applications*, with 99 papers. Thus, it can be said that the publishing sources for these articles are quite varied, especially considering that the number of papers published in the *Lecture Notes In Computer Science* is quite small in relation to the total number of documents analyzed.

TABLE 2. Most cited documents by year.

Year	Article	Cited by
1995	[28]	36
1996	[79]	65
1997	[8]	8
1998	[89]	83
1999	[52]	159
2000	[13]	464
2001	[47]	204
2002	[87]	196
2003	[80]	221
2004	[20]	278
2005	[36]	462
2006	[7]	167
2007	[24]	188
2008	[34]	101
2009	[74]	352
2010	[29]	217
2011	[9]	2069
2012	[3]	82
2013	[30]	125
2014	[48]	154
2015	[63]	237
2016	[11]	147
2017	[17]	137
2018	[26]	137
2019	[51]	26

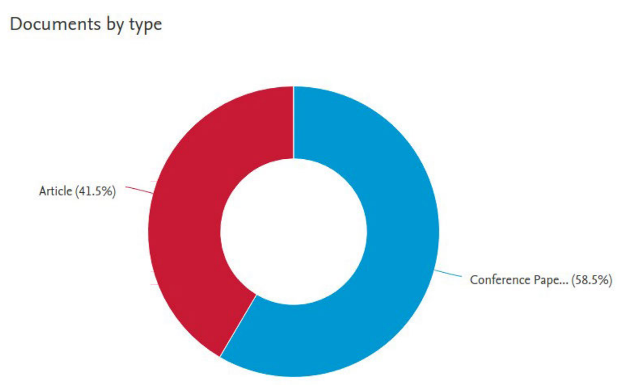


FIGURE 1. Documents by type.

The graph in Figure 1 divides the documents into two types: articles (papers published in journals) and conference papers (papers published in conferences). It is apparent that conference papers are predominant, but the difference between them is relatively small.

The amount of published papers by year is presented in Figure 2, where the curve indicates an exponential increase in documents from 1995 to the present date.

Figure 3 shows the 15 countries with the largest number of publications. China ranks first, with over 350 papers, while the United States ranks second, with almost 300 documents.

K. Hirasawa and S. Mabu are the authors with the largest number of published papers, respectively 27 and 24 articles. There is large difference between second and third place, which has 13 published papers, as shown in Figure 4.

Figure 5 shows the affiliations with the most significant number of published papers, which is surprisingly achieved by a Japanese university. As expected, Chinese universities

TABLE 3. Sources that published most of the documents.

Source	Documents
Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics	180
Expert Systems With Applications	99
Advances In Intelligent Systems And Computing	43
ACM International Conference Proceeding Series	41
Communications In Computer And Information Science	27
European Journal Of Operational Research	24
Applied Soft Computing Journal	23
IEEE IAFE Conference On Computational Intelligence For Financial Engineering Proceedings Cifer	21
Physica A Statistical Mechanics And Its Applications	21
Proceedings Of The International Joint Conference On Neural Networks	18
Information Sciences	16
International Conference On Artificial Intelligence Management Science And Electronic Commerce Aimsec Proceedings	15
IEEE Access	15
Knowledge Based Systems	15
Procedia Computer Science	15

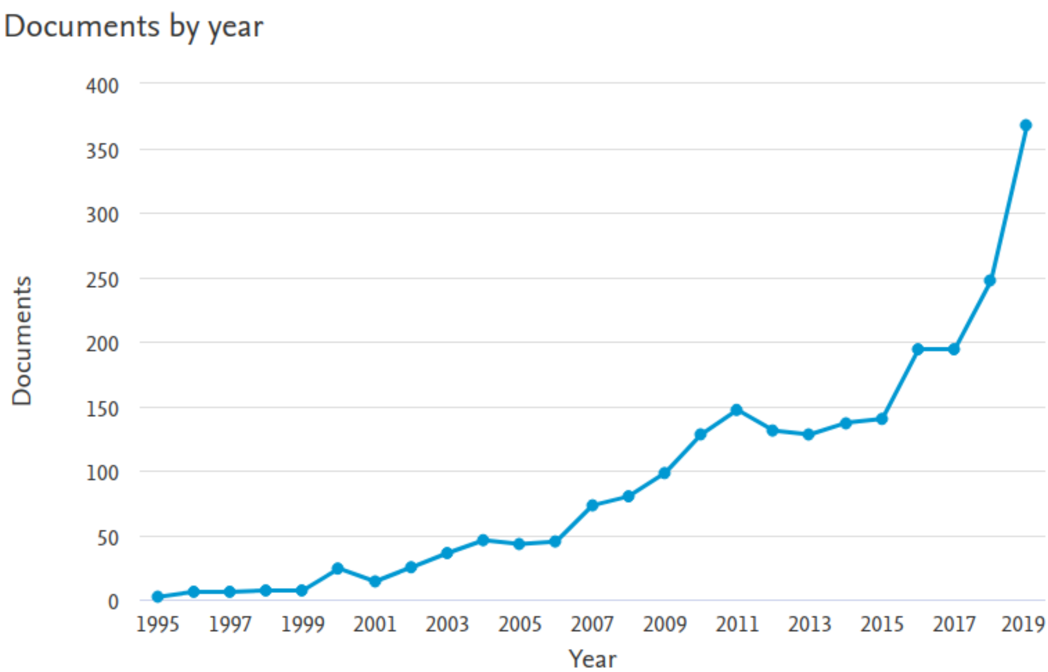


FIGURE 2. Documents by year.

make up most of the 15 affiliations with the largest amounts of publications.

III. PORTFOLIO OPTIMIZATION

Portfolio Optimization, or Portfolio Selection, is a problem that consists of determining a set of financial assets that best suits a particular investor, usually aiming at maximizing profits.

The Modern Portfolio Theory (MPT), created by Markowitz [54], was the first contribution to portfolio optimization models. Markowitz introduced two metrics for evaluating a portfolio's performance: the expected return and the risk. The expected return expresses the idea that an asset that has performed well in the recent past tends to maintain such performance in the future. As a forecast, the risk is the proposed metric to model the return's uncertainty.

The Markowitz model considers the portions of an investor's total budget invested in each asset. Considering n available assets, x_j ($j \in \{1, \ldots, n\}$) is the proportion of the total capital invested in the asset j. Thus, there is a constraint that ensures that the sum of investments must be equal to the capital available for investment, or $\sum_{j=1}^{n} x_j = 1$. In the initial model, short selling was not allowed and, therefore, $x_j \geq 0$, $\forall j \in \{1, \ldots, n\}$.

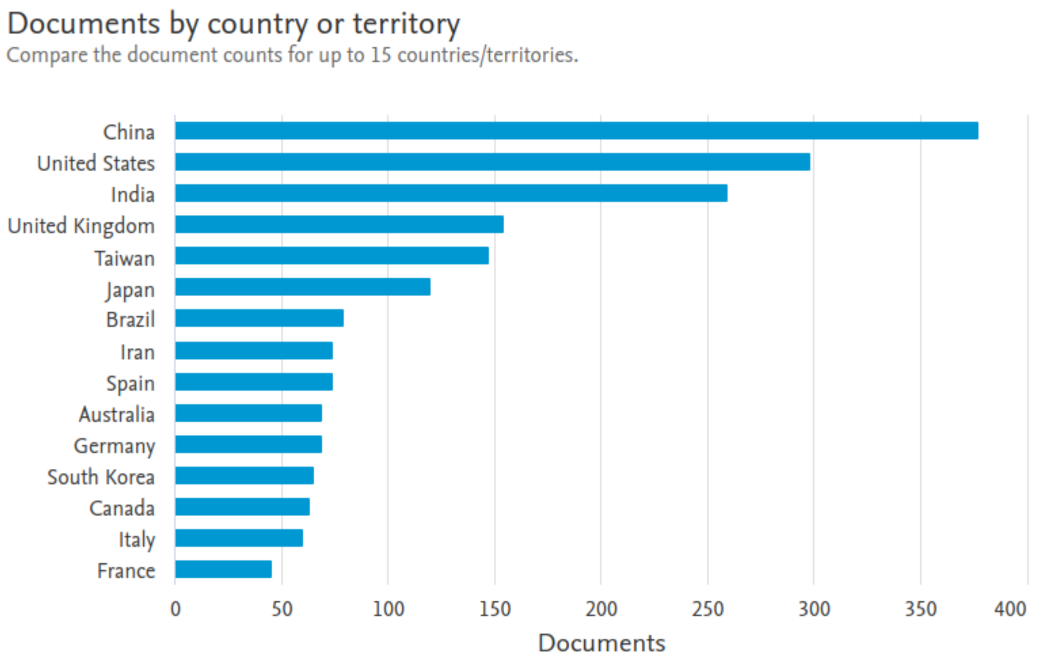


FIGURE 3. **Documents by country.**

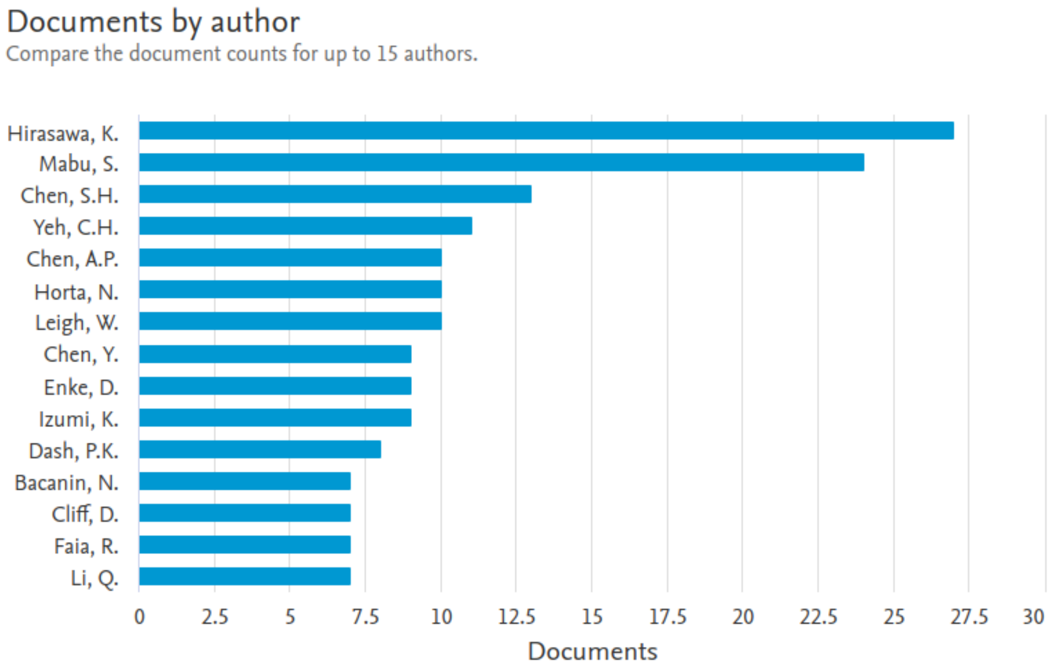


FIGURE 4. **Most prolific authors.**

The aforementioned model uses a random variable R_j, whose average is equal to $\mu_j = E(R_j)$, where $E(R)$ is a function that returns the expected value of the random variable R. This random variable is composed of T different scenarios, in which each scenario t ($t = 1, \ldots, T$) occurs with a probability p_t, so that $\sum_{t=1}^{T} p_t = 1$. In each scenario R_j assumes a value equal to r_{jt}. Thus, the expected value for R_j can be determined by: $\mu_j = \sum_{t=1}^{T} p_t r_{jt}$. In practice, the historical series of rates of return of an asset represents its scenarios. Considering that each scenario occurs with the same probability, the expected return value of an asset j can be defined as follows:

$$\mu_j = \frac{1}{T-1} \sum_{t=2}^{T} r_{jt}$$

where r_j^t is the rate of return of the asset j in the period between $t-1$ and t of a historical series consisting of T quotations of j. Thus, the return R_X of a portfolio X is the

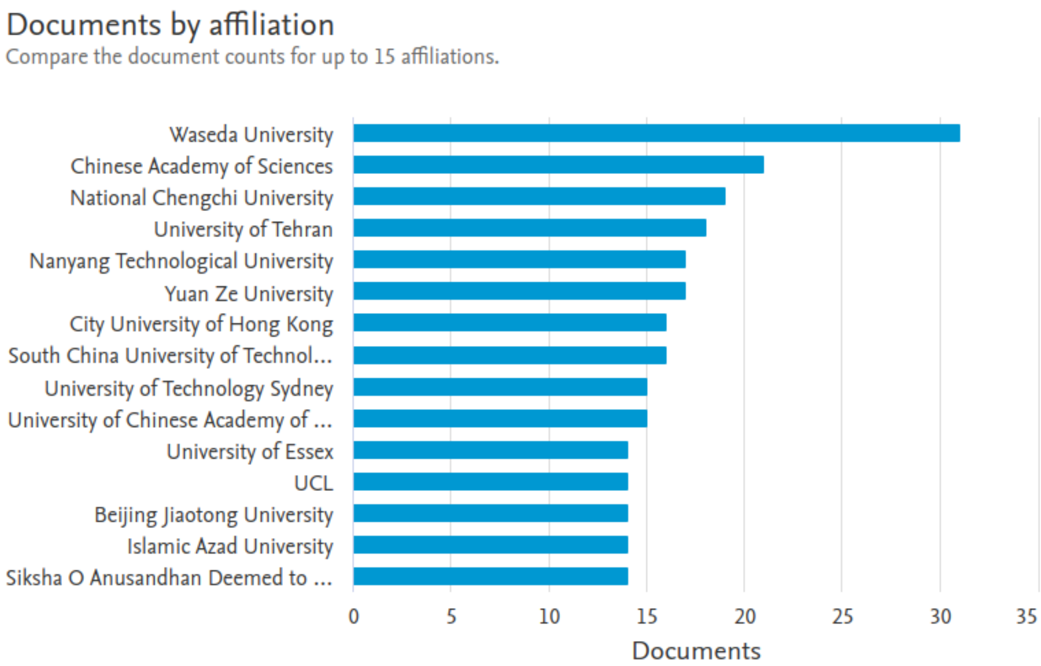


FIGURE 5. **Documents by affiliation.**

weighted average of the returns of the assets that comprise X, that is $R_X = \sum_{j=1}^{n} R_j x_j$, and the expected return of X can be calculated as follows:

$$\mu(X) = E(R_X) = E(\sum_{j=1}^{n} R_j x_j) = \sum_{j=1}^{n} \mu_j x_j. \tag{1}$$

The greatest contribution of the Markowitz model, however, is the introduction to the concept of portfolio risk, defined in its model as the variance of a portfolio's historical returns concerning its expected value:

$$\sigma^2(R_X) = E\{(R_X - E(R_X))^2\}.$$

The portfolio variance is defined based on the pairwise correlation values of the returns of the assets that compose it. Therefore, the diversity of the portfolio tends to reduce its risk [54]. This variance of the portfolio X can be written as follows:

$$\sigma^2(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j.$$

Finally, the classic Markowitz model for portfolio optimization is:

$$\begin{aligned} \min \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \\ \text{subject to: } \sum_{j=1}^{n} \mu_j x_j \geq \mu_0 \\ \sum_{j=1}^{n} x_j = 1 \\ x_j \geq 0 \quad j = 1, \ldots, n \end{aligned}$$

where μ_0 is a lower limit for the portfolio's expected return.

Markowitz's theory has become very widespread and several changes have been made to its original proposal. The use of portfolio variance as a risk measure, for example, has been widely criticized since the variance takes into account both negative and positive deviations. Then, downside risk measures emerged, taking into account only the worst historical returns of the portfolios. Conditional Value at Risk (CVaR) is a downside risk measure widely used because it is a coherent measure [68]. The CVaR (F) function is given by [69]:

$$F_\alpha(X, \zeta) = \zeta + \frac{1}{(1-\alpha)} \sum_{t=1}^{T} p_t [\sum_{j=1}^{n} -(r_{jt} x_j) - \zeta]^+,$$

where p_t is the probability of scenario t to occur, α is the confidence level, and ζ is the VaR$_\alpha$ value, which is the α-quantile of the portfolio losses.

Numerous works have improved these models, creating more risk measures and proposing restrictions that bring them closer to the practical aspects of stock market trading. Several exact, heuristic and hybrid optimization methods have been proposed to solve these portfolio optimization models, which have become increasingly more complex.

Limiting search on Scopus to papers on portfolio optimization, 693 documents were found. Table 4 shows the most cited portfolio optimization papers by year.

Moreover, a portfolio optimization Mean-Absolute Deviation (MAD) model was proposed with the objective to minimize the absolute deviation (risk measure) [79]. The expected return is the mean return of a historical series, and the minimum return constraint defines its lower bound. The model also considers a cardinality restriction limiting the number of assets of the portfolio to an upper bound.

TABLE 4. Most cited portfolio optimization papers by year.

Year	Paper	Cited by
1995		
1996	[79]	67
1997		
1998		
1999	[52]	160
2000	[13]	467
2001		
2002		
2003	[18]	197
2004	[21]	188
2005	[81]	36
2006	[57]	59
2007	[24]	189
2008	[31]	13
2009	[14]	145
2010	[64]	119
2011	[96]	131
2012	[91]	78
2013	[65]	118
2014	[53]	79
2015	[16]	43
2016	[75]	29
2017	[40]	21
2018	[23]	27
2019	[46]	6

Transaction lots constraint ensures that some of the assets can only be traded in an amount that is a multiple of a minimum quantity of shares. The paper employed a hybrid method, which applies a proposed heuristic that finds good weights for assets after an exact method solves the linear relaxed problem.

In [52], a Mean Semi-Absolute Deviation model aimed to minimize the semi-absolute deviation as the risk measure, and the mean return is constrained to a lower bound value. The model also considered the transaction lots constraint, and proves that this constraint makes the model be NP-complete. Regarding the optimization methods, the paper proposed and compared heuristics that transform linear relaxed exact solutions into feasible solutions, which do not violate the transaction lots constraint.

Note that although the first risk measure proposed by Markowitz is variance, some of the first studies on portfolio optimization attempted to use other measures, such as Absolute Deviation or Semi-absolute Deviation. Constraints were also considered in order to bring the model closer to reality.

[13] considered a Mean-Variance model, in which variance is the risk measure to be minimized, and the mean return is limited to a lower bound value. The portfolio cardinality is constrained to a single value. The paper proposed three heuristic methods based on the Genetic Algorithm, Tabu Search, and Simulated Annealing. [18] also considers a Mean-Variance model, but its model considers a variable cardinality constraint, which allows the portfolio cardinality to assume a range of values. This work proposes a Simulated Annealing algorithm to solve the model.

[21] considered a monobjective Mean-Variance model that combines the portfolio variance minimization with the mean return maximization in a single objective function using complementary weighted factors for each objective. Genetic Algorithm, Tabu Search, and Simulated Annealing based methods are proposed to solve the portfolio optimization model.

[81] proposed a model with three objectives: mean return maximization, variance minimization, and Value-at-Risk (VaR) minimization. The proposed heuristic initializes the population with the Randomized Linear Programming (RLP), generates an interim Pareto front with Pareto Sorting Evolutionary Algorithm (PSEA) and Target Objective Genetic Algorithm (TOGA), completes gaps in the Pareto front with TOGA, and stores the result in a repository.

Reference [57] performed hybrid methods by combining proposed heuristics and Genetic Algorithm to solve the monobjective Mean-Variance model. Reference [24] presents a combination of Neural Artificial Networks with heuristics to solve the weighted factors monobjective Mean-Variance model. Reference [31] proposes a multiobjective evolutionary algorithm based on SPES2 to solve the biobjetctive Mean-Variance model.

Reference [14] compared different weighted factors of monobjective models composed by risk minimization and mean return maximization. The models differ by the risk measures applied: variance, semivariance, and absolute deviation. Genetic Algorithms with specific operators and repair methods for each model were performed.

With the increase in the complexity of the models and considering larger number of assets and a larger historical series, the works started to use heuristics instead of exact methods, since these methods solve complex models in polynomial time.

Reference [64] proposed a portfolio optimization model with three objectives: maximize return (mean), minimize risk (variance), and maximize historical returns skewness. For this model, a Multiobjective Particle Swarm Optimization (MOPSO) algorithm was performed.

Reference [96] performed Particle Swarm Optimization (PSO) algorithm with specific operators and repair operators for three different monobjective optimization models: Mean-Variance model with minimum return constraint, Mean-Variance utility function model using weight factors, and Sharpe-ratio maximization portfolio optimization model.

Reference [91] applied Genetic Algorithm and Simulated Annealing to a proposed portfolio optimization monobjective model in which the objective is a utility function formed by variance minimization and portfolio diversity maximization, and the minimum return constraint defines a lower bound for the mean return.

Reference [65] surveyed two biobjective portfolio optimization model, both considering mean return maximization and different risk measures: variance and VaR minimization. The work considered cardinality, transaction lots, and turnover constraints, and found that these methods can be solved by several Multiobjective Evolutionary Algorithms (MOEAs). Finally, the paper compares the models mentioned above to a monobjective Sharpe-ratio model solved by a Genetic Algorithm.

Reference [53] reviewed monobjectives and biobjectives portfolio optimization methods, which aim to minimize risk, subject to a minimum mean return value, and to minimize risk and maximize mean return, respectively. The considered risk measures include absolute deviation, minimum return, Gini's Mean Difference (GMD), and Conditional Value-at-Risk (CVaR). The paper analyzed transaction costs, cardinality, and transaction lots constraints. The methods reviewed can be divided into heuristics: Non-dominated Sorting Genetic Algorithm II (NSGA-II), Pareto Envelope-based Selection Algorithm (PESA) and Strength Pareto Evolutionary Algorithm 2 (SPEA2) for multiobjective models, and Genetic Algorithm (GA) and Threshold Accepting (TA) for the monobjective ones; and exacts: branch-and-bound and branch-and-cut based methods; and hybrids: methods in which heuristics find a good relatively small subset of assets before exact algorithms find optimum solutions for these subsets.

Another trend that was subsequently observed is the use of multiobjective models, since the choice of prioritizing return or risk depends on the profile of a particular investor. In a multiobjective model, a set of non-dominated portfolios are provided so that the best portfolio according to the investor profile can be subsequently chosen.

Reference [16] proposed a utility function monobjective model representing historical portfolio returns as a random fuzzy variable. The utility function combines the random fuzzy variable variance (risk measure) minimization and its mean (expected return) maximization. The model considers cardinality, transaction costs, and turnover constraints. A modified Artificial Bee Colony (ABC) algorithm was performed to solve the model.

Reference [75] proposed a hybrid harmony search and artificial bee colony algorithm to solve the monobjective mean-semi variance portfolio optimization model with minimum return and cardinality constraints. Reference [40] presented specific repair operators for an Artificial Bee Colony algorithm performed to solve the monobjective mean-variance model using a utility function with weighted factors, limited to a single portfolio cardinality value.

Reference [23] surveyed Swarm Intelligence algorithms applied to portfolio optimization models. The analyzed algorithms are PSO (Particle Swarm Optimization), BPO (Business Process Optimization), ACO (Ant Colony Optimization), ABC (Artificial Bee Colony), CSO (Cat Swarm Optimization), FA (Firefly Algorithm), IWO (Improved invasive weed optimization), BA (Bat Algorithm) and FWA (Fireworks Algorithm). These models aim to minimize risk, subject to minimum return, cardinality, transaction costs, and transaction lots constraints. The considered risk measures are variance, variance with skewness, semivariance, mean absolute deviation (MAD), Value-at-Risk (VaR), minimum return, and Conditional Value-at-Risk (CVaR).

Reference [46] proposed an Iterated Local Search (ILS) heuristic in which the local searches are performed using a quadratic programming algorithm to solve a monobjective mean-variance portfolio optimization model subject to minimum return and cardinality constraints.

Recently, [15] introduced a robust multiobjective optimization model based on the mean-variance model and elaborates a multiobjective Particle Swarm Optimization (PSO) algorithm for the specific problem. Reference [76] also use a multiobjective PSO algorithm, but proposes a PSO with ranks to solve the medium-variance model with variable cardinalilty constraints. Reference [43] developed NSGA-II and SPEA2 algorithms with specific operators for three different multiobjective models, which intend to maximize the return and minimize the risk, differing in relation to the risk measure considered: semivariance, CVaR, and a combination of both. Reference [25] proposed a biobjective mean-CVaR model with lots, variable cardinality, and turnover constraints, in addition to an evolutionary algorithm based on the NSGA-II to solve it. Finally, the paper suggested three different decision-making methods for selecting a single portfolio on the Pareto-optimal border, based on a given investor's profile.

It can be observed that recent works tend to use models with two or more objectives and propose heuristics to solve them, since the presence of several objectives increases the complexity of the model's solution by exact methods.

Table 5 presents the main model characteristics from the papers collected from Scopus. These characteristics are the number of objectives for the proposed or studied models in each paper (some of them consider multiple models with different numbers of objectives, such as [65]), the presence of weight factors, which apply weights for different objectives summed in one single objective (x indicates the presence of this characteristic), the use of fuzzy variables in the model, the use of skewness or diversification of the historical series to evaluate the portfolio.

As portfolio risks can be measured in several different ways, Table 6 presents the risks measures considered in each paper, where evaluated models that consider multiple risk measures, while others analyze various models with different risk measures. In this table, V stands for the Variance risk measure, SV for Semi Variance, AD for Absolute Deviation, SAD for Semi Absolute Deviation, VaR for Value-at-Risk, CVaR for Conditional Value-at-Risk, MR for Minimum Return, and GMD for Gini's Mean Difference.

In addition to the objectives, which are usually risks and returns, portfolio optimization models are also characterized by their constraints. Table 7 presents the constraints used in the models of each paper. The main constraints are:

- Minimum return constraint defines a lower limit for the portfolio's expected return.
- Transaction lots constraint ensures that the capital invested in a given asset is a multiple of a minimum number of shares.
- Cardinality constraint limits the number of assets that can make up the portfolio. Fixed cardinality constraint allows only one cardinality value for the portfolio, while variable cardinality constraint enables a range

TABLE 5. Portfolio optimization models.

Paper	Objectives	Weight factors	Fuzzy variables	Skewness	Diversification
[79]	1				
[52]	1				
[13]	1				
[18]	1				
[21]	2	x			
[81]	3				
[57]	1				
[24]	1	x			
[31]	2				
[14]	1	x			
[64]	3			x	
[96]	1	x			
[91]	1	x			x
[65]	1, 2				
[53]	1, 2				
[16]	1	x	x		
[75]	1				
[40]	1	x			
[23]	1	x		x	
[46]	1				

TABLE 6. Portfolio optimization risks.

Paper	V	SV	AD	SAD	VaR	CVaR	MR	GMD
[79]			x					
[52]				x				
[13]	x							
[18]	x							
[21]	x							
[81]	x				x			
[57]	x							
[24]	x							
[31]	x							
[14]	x	x	x					
[64]	x							
[96]	x							
[91]	x							
[65]	x				x			
[53]			x			x	x	x
[16]				x				
[75]		x						
[40]	x							
[23]	x	x	x		x	x	x	
[46]	x							

of integer values for the number of assets in the portfolio.
- Transaction costs are related to the trade of assets in the Stock Market. Some portfolio optimization models consider the constraint that ensures that the total cost of the portfolio, including transaction costs, cannot exceed the capital available for investment.
- Turnover constraint is used in multi-period optimization, and for each period, it defines the cost of a new portfolio based on an existing current portfolio.

Optimization methods can be heuristic, exact, or hybrid when both are used. There is still the possibility of using Machine Learning (ML) algorithms to assist in the optimization process. Multi-Attribute Utility Theory (MAUT) a posteriori methods are often used for multiobjective problems when the selection of only one solution is required. So, Table 8 presents the types of methods performed in each paper.

Portfolio optimization models are becoming increasingly complex, presenting more restrictions and, in some cases, several objectives. This way, there is a tendency to use heuristics to solve them since exact methods cannot solve some more complex models in polynomial time. A more significant number of objectives stems from the growing number of metrics proposed to represent the return and, principally, the risk of a financial portfolio.

IV. STOCK MARKET PREDICTION USING ARTIFICIAL INTELLIGENCE

Stock market prediction or forecasting using historical time series has become a technique widely used by researchers and investors to obtain financial profits in stock trading. These predictions, initially carried out by statistical methods, have been increasingly performed by Artificial Intelligence algorithms. Therefore, AI applied to investments constitutes a

TABLE 7. Portfolio optimization constraints.

Paper	Minimum Return	Transaction Lots	Cardinality	Transaction costs	Turnover
[79]	x	x	x		
[52]	x	x			
[13]	x		x		
[18]	x		x		
[21]					
[81]					
[57]	x		x		
[24]			x		
[31]					
[14]			x		
[64]					
[96]					
[91]	x				
[65]		x	x		x
[53]	x	x	x	x	
[16]	x		x	x	x
[75]	x		x		
[40]			x		
[23]	x	x	x	x	
[46]	x		x		

TABLE 8. Portfolio optimization methods.

Paper	Heuristics	Exact	Hybrid	MAUT	ML
[79]			x		
[52]			x		
[13]	x				
[18]	x				
[21]	x			x	
[81]	x				
[57]	x				
[24]	x				x
[31]	x				
[14]	x				
[64]	x				
[96]	x				
[91]	x				
[65]	x				
[53]	x	x	x		
[16]	x				
[75]	x				
[40]	x				
[23]	x				
[46]			x		

recent research area that has already achieved a large amount of publications.

Since 1965, many researchers have defended the hypothesis of an efficient market [50], which states that the market incorporates all the information that all market participants have and their expectations, so that the price changes are completely random and unpredictable.

In contrast to the efficient market hypothesis, other researchers believe that the market prices fluctuate with a trend. Considering this hypothesis, two schools of market analysis can be regarded as: 1) technical analysis, which defends trends in stock price movements and tries to predict them through historical asset prices; and 2) fundamental analysis, which argues that the socioeconomic context of a company interferes with its future stock price and, therefore, provides information that can be used for forecasting future asset prices [58].

The general framework for an Artificial Intelligence prediction model applied to financial forecasting is presented in Figure 6. The first step is the acquisition of all the necessary data to train and test the predictive model. These data can be treated, transformed, or reduced to remove noisy information and highlight important information. Then, the predictor uses the treated data to train its model, in which its hyperparameters can be optimized in a validation step. Finally, the trained model's performance with tuned hyperparameters needs to be evaluated in a test step.

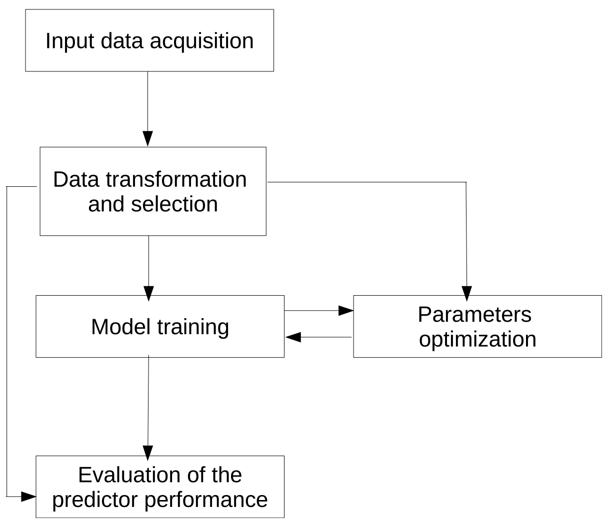


FIGURE 6. Flowchart for general financial forecasting with Artificial Intelligence model predictions.

Limiting search on Scopus to papers on Stock Market forecasts using Machine Learning, 1719 documents were found. Table 9 shows the most cited articles that predict stock market prices or trends using Artificial Intelligence by year.

Reference [28] proposes a Rough Set Theory method that generates rules to assist in Stock Market trading

TABLE 9. Most cited forecasting papers by year.

Year	Paper	Cited by
1995	[28]	36
1996	[72]	30
1997		
1998	[45]	71
1999	[93]	6
2000	[39]	109
2001	[47]	204
2002	[87]	196
2003	[88]	145
2004	[66]	163
2005	[36]	468
2006	[44]	81
2007	[38]	116
2008	[34]	102
2009	[35]	149
2010	[29]	217
2011	[32]	47
2012	[3]	82
2013	[33]	32
2014	[10]	53
2015	[63]	244
2016	[11]	152
2017	[17]	141
2018	[26]	150
2019	[51]	26

actions, which includes buying, selling, and keeping an asset. This method uses fundamentalists indicators as input data. Reference [72] proposed a Probabilistic Neural Network to predict financial prices movement trends using historical assets price series. Three different classes of trends were considered, each one indicating a different action: buying, keeping, and selling a given asset.

Reference [45] presented an Artificial Probabilistic Network (APN) that considers historical prices and fundamentalist indicators as input variables and performs trends classification considering six classes of return levels. Reference [93] employed Support vector machine (SVM), k-Nearest Neighbor Classifier, Probabilistic Neural Network (PNN), Classification and Regression Tree (CART), boosting (Adaboost), and bagging algorithms aiming to perform binary classification of financial assets. Historical prices of the assets were used as an input variable, and the paper's results indicate the better performance of the Boosting algorithm.

Reference [39] used historical prices of assets as input data and predicted price and return with a proposed genetic programming. The paper concludes that asset prices are more predictable than returns. Reference [47] proposed a Genetic Algorithm (GA) integrated with a Fuzzy Neural Network (FNN) model to predict financial trends of assets price movements (considering three different classes of trends) using technical indices as input variables. Reference [87] introduced a Fuzzy grey prediction system that uses historical prices, volume, and fundamentalist index data of assets to predict future prices. Reference [88] proposed a fuzzy rough set system that predicts financial assets future prices using their historical prices and volume data as input.

Reference [66] predicted financial rules that indicate buying and selling signals, performing a proposed genetic programming feed by historical prices and volume data of financial assets. Reference [36] collected fundamentalist indices from financial assets to predict their future price movement trend (in a binary classification) using a Support Vector Machine (SVM) algorithm.

Reference [44] proposed a Genetic Algorithm (GA) that selects the best weights for several classifiers predictions and combines them into a single prediction. The classifiers use technical indicators as input variables and predict financial price movement trends considering four different classes: Bear, Edged-Down, Edged-Up, and Bull. Reference [38] presented three hybrid Artificial Neural Network time series models: NN-EWMA, NN-GARCH, and NN-EGARCH, which uses EWMA, GARCH, and EGARCH to determine input variables before applying the ANN to predict financial assets volatility. The model employs prices, volume, and fundamentalist indicators as input variables, and the results indicate that NN-EGARCH performs best.

Since the first works on prediction of asset trends, Machine Learning algorithms were already being used. It was also common to use heuristics for optimization as Genetic Algorithms, especially for the combination of different classification algorithms.

For instance, [34] combined Support Vector Machine (SVM), Kth Nearest Neighbor (KNN), Back-propagation neural network, decision tree, and logistic regression using a voting committee after performing a wrapper feature selection method. Using prices, volume, and technical indices data to perform a binary assets classification, the paper concludes that voting performs better than single classifiers. Reference [35] first performed a filter-based feature selection for historical prices and technical indicators data. A proposed Self-Organizing Feature Map (SOFM) combined with Support Vector Regression (SVR) was developed to predict future asset prices. SOFM divides training data into several clusters before different SVR models are applied to each cluster. Test data are predicted using the SVR model trained with the most similar cluster.

Reference [29] collected assets historical prices series, which were selected in a stepwise regression analysis (SRA). A Self-organization Map (SOM) neural network was used to divide the training data into clusters, and a fuzzy genetic system was applied to predict assets future prices. Reference [32] proposed a Self-Organizing Map (SOM) combined with Genetic Programming (GP) method. SOM divides training data into several clusters, where each cluster is composed of training, validation, and test data. Validation data select the best GP model for that cluster, and test data evaluate the prediction performance. Prices, volume, and technical indicators are used as input data, and the model predicts the future prices of assets.

Reference [3] proposed the Preprocessed Evolutionary LM Neural Networks (PELMNN). Step-wise Regression Analysis (SRA) for variable selection. A genetic algorithm was used as a global search method to evolve artificial neural networks initial weights, and a Levenberg–Marquardt Back Propagation (LMBP) neural network was trained and used to predict future prices. The input data are technical indices and trading volume. Reference [33] proposed a price forecasting method to perform an iterative feature selection procedure using a backpropagation neural network invalidation data composed of prices, volume, and technical indicators. Finally, a backpropagation neural network was performed on test data to predict the future prices of assets.

Reference [10] selected prices and technical indicators data in a backward elimination method using a Random Forest algorithm. Multiple Random Forest algorithms were used to predict asset's future prices. Final prediction used an average of the predictions of all the Random Forest predictors weighted by their training error. Reference [63] compared ANN, SVM, random forest, and Naive-Bayes algorithms in performing binary classification of assets prices trends. The classifiers are fed by technical indicators data, and a validation step optimizes the classifiers' hyper-parameters. Reference [11] surveyed Stock Market forecasting Machine Learning models. The paper first surveys the preprocessing techniques: normalization, outliers exclusion, clustering, and feature selection. The considered forecasting models are the Artificial Neural Network (ANN) and Support Vector Machine and ensembles methods that combine both and integrate the models mentioned above with heuristics to predict assets future prices or price trends.

Reference [17] compares three different feature selection and transformation methods: Principal Components Analysis (PCA), AutoEncoder (AE), and the Restricted Boltzmann Machine (RBM). Then, a Machine Learning algorithm is performed to predict future asset return. For this, log returns data are collected every five minutes. Reference [26] uses a Deep Learning Long Short Term Memory (LSTM) Neural Network for a binary assets prices trend classification, analyzing a historical series of assets return. Reference [51] proposes a Deep learning model with convolutional and recurrent neurons layers which classifies future assets price trends in three different classes, using prices and volume historical data as input.

A trend that has been increasingly explored is the use of complex techniques for preprocessing the input data, which facilitates the execution of Machine Learning algorithms and increases its accuracy, since the noisy data tend to be eliminated, leaving only the most relevant data.

Recently, there has been an increasing tendency to apply deep neural networks for stock market forecasting. For example, [94] developed a Deep Neural Network (DNN) to classify future trends of asset prices (considering two price directions of the next price). The data set consists of 60 attributes (including returns and technical indicators) of assets belonging to the SPDR S&P 500 ETF between June 2003 and May 2013, with daily frequency. Compared to an Artificial Neural Network (ANN), results show that, although the DNN presents higher accuracy, the ANN provides greater returns and lower risks (variance) in a Stock Market trading simulation. Reference [85] collected open, close, low, and high prices of Yahoo and Microsoft assets from January 2011 to December 2015 and computed five technical indicators as features: momentum, volatility, index momentum, index volatility, stock momentum, stock price volatility. The paper further compares the SVM and LSTM models for the problem of binary classification of stock trends, and the results indicate the better accuracy of the LSTM algorithm. Reference [59] computed ten technical indicators from opening, close, low. The high prices of assets during November 2009 to November 2019 were used to predict prices by applying Decision Tree, Bagging, Random Forest, Adaboost, Gradient Boosting, XGBoost, Artificial Neural Network (ANN), Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) algorithms, concluding that the LSTM algorithm performs better than the others.

Reference [97] proposes a new two-dimensional CNN, which uses a matrix composed of futures, options, opening, closing, high and low prices, and the transaction volume of each asset in a time series of 120 days of data. For that purpose, 5 Taiwanese assets and 5 United States assets are used. The proposed CNN is used to predict trends in the movement of financial asset prices. It considers three different classes: class 1, for days when the return exceeds 1% (upward trend), −1, for days when the return is less than −1% (downward trend) and 0, otherwise (lateral movement). The accuracy of the novel classifier is compared to that of SVM, Neural Network and one-dimensional CNN and the results show that the novel CNN surpasses all other classifiers considering each of the 10 assets.

Table 10 describes the input used by each paper collected on Scopus. Input data can be divided into historical asset prices, historical returns, trading volume, Technical Indices (TI), and Fundamentalist Indices (FI). Feature selection methods were used to remove unimportant features and reduce the number of variables. The period or frequency of sequential data in a time series is a vital data characteristic that can influence the trading strategy.

Table 11 describes the proposed model's output of the different papers. Forecasting financial models can predict different classes of price trends, financial rules to assist in tradings, future asset prices, returns, or volatility.

Other forecasting model characteristics are shown in Table 12, including the uses of Machine Learning (ML) algorithms to make predictions, a heuristic to make the prediction or improve the predictor accuracy or speed, fuzzy systems to improve a predictor or to make trading decisions, clustering of the data before the prediction, ensembles or combinations of different predictors or classifiers, and the application of a validation step to optimize the model's hyperparameters.

For predicting asset prices and their trends, the vast majority of methods found in the literature employed Machine Learning techniques. Regression techniques are commonly used to price prediction, while classification

TABLE 10. Forecasting models input data.

Paper	Price	Return	Volume	TI	FI	Feature Selection	Period
[28]					x		daily
[72]	x						daily
[45]	x				x		daily
[93]	x						daily
[39]	x						daily
[47]				x			daily
[87]	x		x		x		daily
[88]	x		x				daily
[66]	x		x				daily
[36]					x		daily
[44]				x			daily
[38]	x		x		x		daily
[34]	x		x	x			daily
[35]	x			x			daily
[29]	x						daily
[32]	x		x	x			daily
[3]			x	x		x	daily
[33]	x		x	x		x	daily
[10]	x			x		x	daily
[63]				x			daily
[11]	x		x	x	x	x	daily
[17]		x				x	5 minutes
[26]		x					daily
[51]	x		x				daily

TABLE 11. Forecasting models output variables.

Paper	Classes	Rules	Price	Return	Volatility
[28]		x			
[72]	3				
[45]	6				
[93]	2				
[39]			x	x	
[47]	3				
[87]			x		
[88]			x		
[66]		x			
[36]	2				
[44]	4				
[38]					x
[34]	2				
[35]			x		
[29]			x		
[32]			x		
[3]			x		
[33]			x		
[10]			x		
[63]	2				
[11]	2		x		
[17]				x	
[26]	2				
[51]	3				

techniques are frequently used to predict trends in asset price movements. An increasing amount of data used as inputs or features was identified, which implies the need to use methods to select and preprocess this input data to filter only the most essential information. Deep Learning methods are another trend showing promising results in recent works, despite its great computational complexity, which makes selecting and preprocessing data even more important.

V. FINANCIAL SENTIMENT ANALYSIS

Sentiment analysis is the field of study that analyzes people's feelings and moods towards an entity, such as a product to evaluate whether the opinions are negative or positive and how negative or positive they are. To perform sentiment analysis, it is first necessary to collect a large number of texts, such as those extracted from social media or news websites, written in a natural language. Thus, the application of natural language processing methods is also necessary [49]. Sentiment analysis has been increasingly applied in several areas of knowledge, such as computational finance.

In finance, several researchers test the effect of news and opinions on future asset prices, putting in check the hypothesis of the efficient market, based on the idea that news and opinions guide investors, creating trends in market prices and, thus, providing possibilities for an investor to obtain profits in stock market tradings.

Limiting the search on Scopus to papers on financial sentiment analysis, 289 documents were found. The most cited sentiment analysis papers by year are present in Table 13.

Reference [55] collected press related to stocks in NYSE or NASDAQ-AMEX, considering only press releases of companies with a turnover higher than US$5, 000, 000 per day. Text reprocessing removes stopwords, numbers, and less meaningful terms according to the TFxIDF measure. These press releases are classified as "Good News," "Bad News," and "No Movers" (for neutral news). Good news provokes a rise of 3% on the stock price within at least 60 minutes after the press release and increase the price average in at least 1% during this interval. Bad news provokes a drop of 3% in the stock price within at least 60 minutes after the press release and decreases the price average in at least 1% during this interval. The other news is classified as "No Movers." An SVM model was applied to predict stock prices using the news sentiment and historical price series. Results show a recall of approximately 60% and a better cumulative return and average return per trade than the random trader in a stock market simulation.

TABLE 12. Forecasting models characteristics.

Paper	ML	Heuristic	Fuzzy	Clustering	Ensembles	Validation
[28]		x				
[72]	x					
[45]	x					
[93]	x				x	
[39]		x				
[47]	x	x	x		x	
[87]			x			
[88]			x			
[66]		x				
[36]	x					
[44]	x	x			x	
[38]	x					
[34]	x				x	
[35]	x			x		
[29]	x	x	x	x		
[32]		x		x		x
[3]	x	x				
[33]	x					
[10]	x				x	
[63]	x				x	x
[11]	x			x	x	
[17]	x					
[26]	x					
[51]	x					

TABLE 13. Most cited sentiment analysis papers by year.

Year	Paper	Cited by
1995		
1996		
1997		
1998		
1999		
2000		
2001		
2002		
2003		
2004	[55]	92
2005		
2006	[83]	4
2007	[84]	6
2008		
2009	[74]	358
2010	[73]	33
2011	[9]	2100
2012	[86]	48
2013	[78]	50
2014	[48]	158
2015	[60]	164
2016	[61]	49
2017	[62]	56
2018	[5]	12
2019	[2]	5

Reference [83] analyzed 77,256 analysts reports from Thomson Financial Web service related to the companies in the Tokyo Stock Exchange from January 1, 2001 to March 31, 2003. 12 keywords were extracted from the obtained reports' title and classified into three classes: Good, Bad, and Neutral News. Change of monthly consensus earnings estimate for the next fiscal year (CESFY1) is used as a numerical indicator for future stock performance. A Machine Learning model uses a binary classification label according to monthly earning forecast changes: upward or downward revision to predict stock movements. The proposed model performance indicates that the sentiment is related to stock prices 20 days before and 20 days after the press.

Reference [84] extracted specific news that has a high possibility to affect the stock prices from news data offered from JIJI Press about companies in the Tokyo stock exchange market between August 10, 2006 and November 24, 2006. News is classified into three classes: "Good News," "Bad News" and "Neutral News" using Naive Bayes, which presents 78% accuracy. Results show that the average return for 30 days prior and 30 days after the news varied for these different categories, and the average daily return was lower before and after Bad News and greater before and after Good News.

Reference [74] proposed four stock price prediction models with different features. The first model applies linear regression using 60-minute stock quotations before given news; the second one consists of an SVM that uses only extracted article terms for its prediction; the third one is an SVM that uses extracted article terms and the stock price at the time the article was released; and the fourth one is an SVM that uses extracted terms and a regressed estimate provided by the linear regression model of the stock price 20 minutes after the news. The paper concludes that the third model performs the best. Reference [73] compared the methodology proposed by [74] to the top 10 quant funds operating for a full year at the time of the study. The paper concludes that the proposed methodology provides the fourth-best return among the top 10 and the best return when compared with only the quant funds for companies in S&P 500.

In relation to the works that perform analysis of feelings, it is observed that the oldest studies used news about a certain asset to assign a mood to it. A disadvantage of using news is that it tends to be more neutral, which makes it difficult to differentiate between good and bad ones, in addition to the smaller amount, in relation to comments on social media.

Reference [9] employed lexicon dictionaries to perform sentiment analysis to public tweets recorded from February 28 to December 19, 2008 from stocks in DJIA. After stopwords and punctuation removal, a one-dimension public mood time series was generated by OpinionFinder, which classifies text moods as positive or negative. Besides, seven dimensions of the public mood time series were generated by GPOMS, each representing a different aspect of the public's mood on a given day: Calm, Alert, Sure, Vital, Kind, and Happy. Bivariate Granger causality analysis showed that only the GPOMS's Calm dimension correlates (linear) with the stock prices series. Self-Organizing Fuzzy Neural Network was proposed to predict stock movements using three previous historical daily prices and permutations of the seven mood indicators as features. Self-Organizing Fuzzy Neural Network predictions match Granger analysis results, showing that the calm indicator increases the prediction accuracy, but the other indicators do not improve the prediction using only the stock price series.

Reference [86] collected six security companies' quarterly and annual reports and their ROEs time series and selected terms that occurred three times or more. The paper performed future asset prices forecasting using ARIMA and SVR, which uses historical return series, and historical term series from reports, respectively. Results reveal that the hybrid model, combining ARIMA and SVR, presents the best forecasting accuracy.

Reference [78] used a data set of 1,600,000 (800,000 positive and 800,000 negative) tweets collected and labeled by Stanford University. The preprocessing step removed usernames and links, identified occurrences of more than two letters in a word and changed it to only one letter, drew explicit negation words, exclamation, and question marks, besides performing text tokenization, removal of stopwords, stemming, N-gram construction (size 2), and removed words that occurred only once. Two sentiment analysis models were applied. The first classifies the documents into two classes: negative and positive, while the second classifies them into three categories: negative, neutral, and positive. Sentiment analysis presents an accuracy close to 80%. The study then analyzed a correlation between 152,572 tweets discussing stock relevant information about eight companies in nine months in 2011 and the stock closing prices of these eight companies for the same period. A statistical hypothesis test for stationary time series was performed to determine the linear correlation level between the sentiment and stock closing price and whether one contains predictive information about the other (Granger causality analysis). Results indicate that tweets' sentiment can predict stock price movements for several assets, and the introduced neutral class can improve the correlation between the opinionated tweets and the stock closing price in certain situations.

Reference [48] examined a news archive from FINET, containing both company-specific and market-related news from January 2003 to March 2008. The news was classified by Harvard IV-4 sentiment dictionary (HVD), considering 15 sentiment dimensions, and Loughran–McDonald financial sentiment dictionary (LMD), considering six sentiment dimensions. An SVM model was developed to predict stock movement using historic daily open-to-close price return as a feature, in addition to the news sentiment moods values. For the SVM classification, the stocks were labeled into three classes according to return value: positive, neutral, and negative. The paper concludes that sentiment analysis helps to improve prediction accuracy.

Reference [60] proposed different methods to classify untagged messages in five classes: Strong Buy, Buy, Hold, Sell, and Strong Sell. These messages were taken from 18 message boards of the 18 stocks from Yahoo Finance Message Board for one year (July 23, 2012 to July 19, 2013). Historical daily adjusted close prices extracted from Yahoo Finance for 18 stocks were also used as features for a stock movement prediction performed by an SVM algorithm, in which its labels represent actual price movement (up or down). Results indicate that the incorporation of the sentiment analysis improves the prediction accuracy.

Reference [61] employed all tweets containing cashtags of all stocks traded in US stock markets from December 22, 2012 to March 27, 2015, before the Stanford CoreNLP tool was applied to execute common natural language processing methods, including tokenization, Part Of Speech (POS) tagging, and lemmatization. Sentiment analysis classification considers three sentiment classes: "bullish," "bearish," and "neutral." The work measured the correlation between Twitter sentiment indicators and two popular survey sentiment indicators: the American Association of Individual Investors (AAII) and Investors Intelligence (II). A strong correlation between them indicates that the microblogging sentiment indicator can provide important information.

Reference [62] collected a total of 2,50,000 tweets about Microsoft from August 31, 2015 to August 25, 2016, extracted from Twitter API, in addition to stock opening and closing prices of Microsoft in the same period obtained from Yahoo! Finance. Tweet preprocessing included tokenization, stopwords removal, and regex matching for removing special characters. Sentiment analysis considered three different classes for each tweet: positive, neutral, and negative. A total of 3,216 tweets were examined, labeled, then used to train a Random Forest classification model. The trained model predicts other tweets' sentiments. After performing sentiment analysis, a classification model used positive, neutral, and negative tweets in 3 days as features to perform a binary stock movement classification. Results show around 70% accuracy for sentiment classification and around 70% accuracy for stock movement prediction.

Reference [5] retrieved around 300,000 tweets about Apple from StockTwits during 2010-2017, each tweet composed by its content, date, and user sentiment. Tweet preprocessing included tokenization and removal of stopwords and Twitter symbols. An SVM model was trained for sentiment prediction, in which daily sentiment is positive if there are more positive tweets than negative ones and is negative if otherwise. Besides, Apple's historical price data were extracted from Yahoo Finance from 2010 to 2017 and used for an SVM classifier that predicts binary class stock movement (up or down). For sentiment prediction, the achieved test accuracy was 63.5%, 75.3% recall, and 76.8% precision, and for stock movement prediction, the completed test accuracy was 76.68%, 100% recall, and 69.5% precision.

Reference [2] assigned daily sentiment scores for the DJIA market by accumulating high-frequency sentiment scores of the DJIA's constituents obtained from the TRNA dataset from January 2006 to October 2012. DJIA's constituents were tagged according to three classes: positive, neutral, or negative (and a probability associated with each one). Then, the daily sentiment was computed as the average sentiment prediction weighted by probabilities of each of these predictions to be corrected. The paper applied linear and quantile regression to these daily scores, analyzing its correlation with the Thomson Tick History database's stock prices. The regressions considered daily sentiments up to 5 days before a given price and that price. Results demonstrate that daily financial news sentiment can predict prices, as they show a significant correlation.

Recently, [56] considered news articles for the S&P 500 companies from February 2013 to March 2017 from international daily news websites, comprising a total of 265463 articles, in addition to daily closing stock prices of these assets for the same interval. The paper developed an ARIMA model for price regression, considering only the stock price data. The Facebook Prophet algorithm was used to predict future stock prices using historical closing prices. Reference [56] also proposed three Recurrent Neural Network Long Short Term Memory (RNN LSTM) methods for stock price prediction. The first method uses only the price data as features; the second uses historical closing prices and the textual polarity (negative or positive) for each asset provided by the natural language toolkit (NLTK); and the third uses the prices and textual data as features. The paper concludes that there is a strong relationship between stock prices and financial news articles, as the RNN LSTM models that use textual data or sentiment information perform better than models that use only the prices data. Reference [90] first captured the semantic information of stock-related tweets texts by applying LSTM models that produce a textual representation for each set of texts about a particular asset on a given day. Subsequently, another LSTM model followed by a regression MLP was used to predict future asset prices. Reference [41] proposed using the Convolutional Neural Network (CNN) algorithm to extract features from the set of words formed by the text of each tweet. The paper also presented the optimization of the generated features using the Cuckoo Search (CS) Algorithm and, finally, the classification of the texts' polarity (positive or negative) using a Neural Network algorithm. Results showed that this proposed approach surpasses previous approaches found in literature, considering the accuracy measure.

Reference [70] proposed a Sentiment-aware volatility forecasting (SAVING) method. The sentiment analysis considered social media messages on StockTwits for 10 US stocks from August 14, 2017 to August 22, 2018. The work assessed the polarity of feeling for a given asset taking into account the intensity and quantity of messages about the asset. This polarity was coupled to the volatility prediction models as variables or features. For this, the paper used the Recurrent Neural Network (RNN) algorithm. The proposed SAVING model was compared with volatility prediction models that do not use sentiment analysis: GARCH, EGARCH, TARCH, GJR, GP-vol, VRNN, NSVM and LSTM (Long Short Term Memory) models. A t-test compared SAVING with each other method and results indicate that SAVING statistically outperforms all other methods, except EGARCH, TARCH and GJR, in addition to not being dominated by any other method.

Information about companies can be found on reported news, available on several websites, or comments and opinions given by social media users. Table 14 informs the data source used by each paper.

TABLE 14. Text source.

Paper	News	Social media comments
[55]	x	
[83]	x	
[84]	x	
[74]	x	
[73]	x	
[9]		x
[86]	x	
[78]		x
[48]	x	
[60]		x
[61]		x
[62]		x
[5]		x
[2]	x	

Table 15 specifies how the papers attribute the sentiment prediction for each financial stock or index using texts about its companies. Some works developed Machine Learning (ML) models to predict the asset's sentiment using texts collected from the news or social media comments. In contrast, other papers utilized lexicon dictionaries that assign sentiment values based on the collected texts. Other documents employed data sets that contain the sentiment value associated with the text.

Research in stock market sentiment analysis usually tests the impact of news or opinions on the stock prices, challenging the efficient market hypothesis. For this, some papers analyzed the correlation between the sentiment value for a given company and its stock price in the near future (a few days after the news or comment publication). Other papers used these sentiment values to predict stock prices or movement, intending to obtain profits in stock market

TABLE 15. Sentiment analysis method.

Paper	ML	Dictionaries	Data including sentiment
[55]	x		
[83]	x		
[84]	x		
[74]	x		
[73]	x		
[9]		x	
[86]	x		
[78]	x		x
[48]		x	
[60]	x		
[61]	x		
[62]	x		
[5]	x		
[2]			x

tradings. Table 16 shows the application of the sentiment analysis by each paper.

It is noted that the first works on sentiment analysis applied to finance use news as input data. However, there is a tendency to use comments on social networks, which considerably increases this data's total size. Thus, the pre-processing of these data and selecting the most important terms becomes a necessary task in this approach. The sentiment analysis applied to financial investments is commonly used, in practice, to decide whether to buy or sell an asset under study and, therefore, is frequently combined with the historical analysis of prices of these assets. In this combination, the predicted sentiment is used as one of the Machine Learning algorithm features to predict the asset's price or trend of movement.

VI. COMBINATIONS

Combination papers integrate two or more approaches discussed in this work, usually aiming to improve the performance of a Machine Learning model or increase profits provided by trading simulations in the stock market. Although the vast majority of papers used only one approach, many articles indicate better combined methods, which are more recent works.

Reference [27] used historical stock indicators and news to predict stock market movements, analyzing 72 S&P 500 assets from 15 September 2006 to 31 August 2007, collecting historical prices, trading volumes, best bid and ask prices, in addition to 10 technical indicators. Sentiment analysis classifies each news document as negative or positive, and a linear regression is performed to assign values that indicate how positive or negative a news item is. A classification model was used to predict a particular asset's chances exceeding a return of 1% over the next period. For this, news polarity, prices, and technical indicators were selected and used as features. Results of the paper show that integrating market data with textual data contributes to improving the performance of the model compared to using only market data and that using more advanced textual data representations further improves the model's prediction accuracy.

Reference [19] applied portfolio optimization methods after performing financial time series classification for stock price movement prediction. For the classification method, 30 attributes were considered among different log-returns and technical indicators obtained from daily data of 115 assets that participated in B3 between January 2004 and December 2016. For each day, t, assets that reach a minimum desired return in the next day $(t+1)$ are labeled as an up-trended asset and the others as down-trended. The classification algorithms are SVMs with different hyperparameters values and a voting ensemble method considering all SVM classifiers. Portfolio optimization considers the monobjective mean-CVaR model with variable cardinality constraint. The precision metric was selected to evaluate the classifiers, and results show that the combination of all classifiers does not exceed the best of the individual classifiers. From trading simulations, the paper concluded that the combination of financial series classification and portfolio optimization surpasses each of the single approaches, as statistical tests indicate better daily returns when using the proposed combination.

Reference [71] combined sentiment analysis and network analysis and tested the correlation between the network sentiment and stock market prices. For this, tweets from three official certificated Twitter accounts (StockTwits, FinancialTimes, and MarketWatch) were collected in a first step. Then, tweets from all the three groups' followers were collected and combined into a single group or community. Tweets from this community were collected from January to September 2015, for 2,898,756 users and 775,928,121 tweets. In addition to this data set, another was formed by a trusted network, including tweets of users tagged by all followers mentioned above. Both data sets were limited to tweets related to the eight firms with the largest number of tweets, considering both. After collecting these eight companies' historical returns from 1 January 2015 to 31 August 2015, linear correlation methods were conducted. Results show a higher correlation when using the trusted network than using only its followers' three official sources.

Reference [95] used the opinion messages extracted from StockTwits, the historical closing price and the daily trading volumes of stocks from the Quandl API, and the market capitalization data from Yahoo! Finance, from 14 August 2017 to 16 November 2017. The paper proposed the use of sentiment analysis to create market views, which were later used in the Black-Litterman asset selection model. For sentiment analysis, the extracted documents were analyzed and classified using a Recurrent Neural Network (RNN) composed of an Evolutionary Clustering Method (ECM) and a Long Short Term Memory (LSTM) network. Results indicate that the proposed model surpasses other Black-Litterman models that do not consider sentiment analysis.

Reference [92] suggested the use of public mood in the selection of financial portfolios. For that, classification

TABLE 16. Stock market sentiments application.

Paper	Correlation analysis	Price prediction	Stock movement prediction
[55]			x
[83]			x
[84]	x		
[74]		x	
[73]		x	
[9]	x		x
[86]		x	
[78]	x		
[48]			x
[60]			x
[61]	x		
[62]			x
[5]			x
[2]	x		

models were proposed and the sentiment information and the historical returns of each asset were used to assign a weight to each asset. A classification was performed for each asset and its weight was determined by the normalization of the score function predicted for each one. The paper employed 15 different stocks from 24 January 2012 to 2 June 2017. Daily financial data from the Quandl API and sentiment data from the StockFluence API were extracted. Results indicate that using sentiment analysis, together with financial data, provides better portfolios than using only financial data.

Reference [67] combined sentiment analysis and stock movement prediction based on historical prices and technical indicators. Sentiment analysis was applied on posts and documents of the 51 assets from the Sina stock forum and Eastmoney stock forum between June 17, 2014 and June 7, 2016. Each sentence of a paper was assigned to a polarity based on HowNet and Chinese Sentiment Analysis Ontology Base. Daily sentiment measure was computed based on the number of negative and positive documents. This measure assumes a value ranging from -1 to 1, where -1 is assigned to days with only negative papers and 1 to days with only positive documents. The article also proposes an SVM binary classification method that predicts stock price movement (up or down) for the SSE 50 Index using eight features based on the sentiment analysis for each of the 51 assets, in addition to the historical opening price, closing price, high for the day, low for the day, trading volume in the number of shares, trading volume in RMB, change in RMB, and change in percentage data of the Index. Results show that the model achieves accuracy ranges of 63%-72% and 80%-90% by using only historical prices and incorporating sentiment analysis features, respectively.

Reference [4] applied deep reinforcement learning algorithm and sentiment analysis in forecasting stock market trends. The proposed model considers stock information, capital, stock assets, and predicted stock trend as an environment for the reinforcement learning model, while the agent's actions are buying, selling, and holding assets. The trend prediction was made using a deep neural network composed of a convolutional layer and LSTM layer, which considers news (represented by text embedding) and historical asset prices as features.

VII. CONCLUSION

This paper consists of a systematic review of the literature on Artificial Intelligence applied to investments in the stock market. Articles were selected from the Scopus website, where documents with the highest number of citations were considered most relevant. The papers were then divided into portfolio optimization, stock market prediction using Artificial Intelligence, financial sentiment analysis, and combination papers involving two or more fields.

The overview presents general information, such as the most cited papers in general and the most cited papers per year, which allows the identification of periods of increased interest in Artificial Intelligence for investments, which, unsurprisingly, coincides with significant technological improvement and popularization of the computer. Another essential piece of information is the number of papers per year, which shows an exponential increase in the number of documents from 1995 to 2019, indicating this to be a recent area that has increasingly attracted research attention.

For each area, primary (most-cited) papers for each year were analyzed, for the period between 1995 and 2019. Besides, recent articles were found to include the state-of-the-art findings for each area, in addition to indicating possible directions for future works. Tables were created for the apparent separation of the different characteristics of the methods and models proposed in each paper.

Finally, the combinations of approaches presented in the papers indicates superiority over single methods, although the number of combination papers in the literature is still small.

A. CHALLENGES, GAPS AND FUTURE INSIGHTS

The biggest challenge in reviewing the papers is the creation of queries for Scopus that exclude irrelevant documents but, at the same time, include the largest amount of relevant documents possible. The second major difficulty is to analyze the large number of documents returned by the query. Thus, this work examines only the most essential documents (those with the highest number of citations). However, natural language processing and data mining methods can be applied in future works to analyze all returned papers.

For the optimization of financial portfolios, it is possible to verify an increasing trend in using multiobjective models and heuristic methods. Thus, models are becoming increasingly complex, and fast methods are desirable. However, the number of papers that compare these more complex models with more basic monobjective models and exact methods with heuristics is still minimal.

Regarding forecasts using historical data of assets, Machine Learning methods are increasingly being used as of recent. Specifically, a significant increase in Deep Learning methods, which has produced promising results. Meanwhile, there are wide variations in features used in different works, including technical and fundamentalist indicators. Yet, limited research has been conducted to analyze each of these feature's contribution to the performance of the methods.

The first works on sentiments analysis applied to financial investment employed the news as the object of study. Due to the popularization of social networks, a large number of comments about financial assets and their companies are available on the Internet, which is becoming the most used source of information in this type of analysis. However, very few papers consider using both sources (news and comments on social networks).

REFERENCES

[1] S. Aghabozorgi and Y. W. Teh, "Stock market co-movement assessment using a three-phase clustering method," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1301–1314, Mar. 2014.

[2] D. E. Allen, M. McAleer, and A. K. Singh, "Daily market news sentiment and stock prices," *Appl. Econ.*, vol. 51, no. 30, pp. 3212–3235, Jun. 2019.

[3] S. Asadi, E. Hadavandi, F. Mehmanpazir, and M. M. Nakhostin, "Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction," *Knowl.-Based Syst.*, vol. 35, pp. 245–258, Nov. 2012.

[4] A. Azhikodan, A. G. K. Bhat, and M. V. Jadhav, "Stock trading bot using deep reinforcement learning," in *Innovations in Computer Science and Engineering*. Singapore: Springer, 2019, pp. 41–49.

[5] R. Batra and S. M. Daudpota, "Integrating StockTwits with sentiment analysis for better prediction of stock price movement," in *Proc. Int. Conf. Comput., Math. Eng. Technol.*, Jan. 2018, pp. 1–5.

[6] V. Boginski, S. Butenko, and P. M. Pardalos, "Statistical analysis of financial networks," *Comput. Statist. Data Anal.*, vol. 48, no. 2, pp. 431–443, Feb. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167947304000258

[7] V. Boginski, S. Butenko, and P. M. Pardalos, "Mining market data: A network approach," *Comput. Oper. Res.*, vol. 33, no. 11, pp. 3171–3184, Nov. 2006.

[8] P. J. Bolland and J. T. Connor, "A constrained neural network Kalman filter for price estimation in high frequency financial data," *Int. J. Neural Syst.*, vol. 8, no. 4, pp. 399–415, Aug. 1997.

[9] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, Mar. 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S187775031100007X

[10] A. Booth, E. Gerding, and F. McGroarty, "Automated trading with performance weighted random forests and seasonality," *Expert Syst. Appl.*, vol. 41, no. 8, pp. 3651–3661, Jun. 2014.

[11] R. C. Cavalcante, R. C. Brasileiro, V. L. F. Souza, J. P. Nobrega, and A. L. I. Oliveira, "Computational intelligence and financial markets: A survey and future directions," *Expert Syst. Appl.*, vol. 55, pp. 194–211, Aug. 2016.

[12] M. Chang. (2018). *How A.I. Traders Will Dominate Hedge Fund Industry*. [Online]. Available: https://www.youtube.com/watch?v=lzaBbQKUtAA

[13] T.-J. Chang, N. Meade, J. E. Beasley, and Y. M. Sharaiha, "Heuristics for cardinality constrained portfolio optimisation," *Comput. Oper. Res.*, vol. 27, no. 13, pp. 1271–1302, Nov. 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S030505489900074X

[14] T.-J. Chang, S.-C. Yang, and K.-J. Chang, "Portfolio optimization problems in different risk measures using genetic algorithm," *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10529–10537, Sep. 2009.

[15] C. Chen and Y. Wei, "Robust multiobjective portfolio optimization: A set order relations approach," *J. Combinat. Optim.*, vol. 38, no. 1, pp. 21–49, Jul. 2019, doi: 10.1007/s10878-018-0364-9.

[16] W. Chen, "Artificial bee colony algorithm for constrained possibilistic portfolio optimization problem," *Phys. A, Stat. Mech. Appl.*, vol. 429, pp. 125–139, Jul. 2015.

[17] E. Chong, C. Han, and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," *Expert Syst. Appl.*, vol. 83, pp. 187–205, Oct. 2017.

[18] Y. Crama and M. Schyns, "Simulated annealing for complex portfolio selection problems," *Eur. J. Oper. Res.*, vol. 150, no. 3, pp. 546–571, Nov. 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221702007841

[19] F. D. Paiva, R. T. N. Cardoso, G. P. Hanaoka, and W. M. Duarte, "Decision-making for financial trading: A fusion approach of machine learning and portfolio selection," *Expert Syst. Appl.*, vol. 115, pp. 635–655, Jan. 2018.

[20] K. Doerner, W. J. Gutjahr, R. F. Hartl, C. Strauss, and C. Stummer, "Pareto ant colony optimization: A Metaheuristic approach to multiobjective portfolio selection," *Ann. Oper. Res.*, vol. 131, nos. 1–4, pp. 79–99, Oct. 2004, doi: 10.1023/B:ANOR.0000039513.99038.c6.

[21] M. Ehrgott, K. Klamroth, and C. Schwehm, "An MCDM approach to portfolio optimization," *Eur. J. Oper. Res.*, vol. 155, no. 3, pp. 752–770, Jun. 2004.

[22] D. Enke and S. Thawornwong, "The use of data mining and neural networks for forecasting stock market returns," *Expert Syst. Appl.*, vol. 29, no. 4, pp. 927–940, Nov. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417405001156

[23] O. Ertenlice and C. B. Kalayci, "A survey of swarm intelligence for portfolio optimization: Algorithms and applications," *Swarm Evol. Comput.*, vol. 39, pp. 36–52, Apr. 2018.

[24] A. Fernández and S. Gómez, "Portfolio selection using neural networks," *Comput. Oper. Res.*, vol. 34, no. 4, pp. 1177–1191, Apr. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0305054805002042

[25] G. H. M. Mendonça, F. G. D. C. Ferreira, R. T. N. Cardoso, and F. V. C. Martins, "Multi-attribute decision making applied to financial portfolio optimization problem," *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113527. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417420303511

[26] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, Oct. 2018.

[27] T. Geva and J. Zahavi, "Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news," *Decis. Support Syst.*, vol. 57, no. 1, pp. 212–223, Jan. 2014.

[28] R. H. Golan and W. Ziarko, "A methodology for stock market analysis utilizing rough set theory," in *Proc. Conf. Comput. Intell. Financial Eng.*, Apr. 1995, pp. 32–40.

[29] E. Hadavandi, H. Shavandi, and A. Ghanbari, "Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting," *Knowl.-Based Syst.*, vol. 23, no. 8, pp. 800–808, Dec. 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705110000857

[30] M. Hagenau, M. Liebmann, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decis. Support Syst.*, vol. 55, no. 3, pp. 685–697, Jun. 2013.

[31] G. Hassan and C. D. Clack, "Multiobjective robustness for portfolio optimization in volatile environments," in *Proc. 10th Annu. Conf. Genetic Evol. Comput.*, 2008, pp. 1507–1514.

[32] C.-M. Hsu, "A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming," *Expert Syst. Appl.*, vol. 38, no. 11, pp. 14026–14036, May 2011.

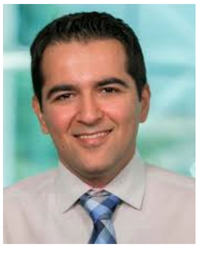
[33] C.-M. Hsu, "A hybrid procedure with feature selection for resolving stock/futures price forecasting problems," *Neural Comput. Appl.*, vol. 22, nos. 3–4, pp. 651–671, Mar. 2013.

[34] C. Huang, D. Yang, and Y. Chuang, "Application of wrapper approach and composite classifier to the stock trend prediction," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2870–2878, May 2008.

[35] C.-L. Huang and C.-Y. Tsai, "A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1529–1539, Mar. 2009.
[36] W. Huang, Y. Nakamori, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," *Comput. Oper. Res.*, vol. 32, no. 10, pp. 2513–2522, Oct. 2005.
[37] W. Q. Huang, X. T. Zhuang, and S. Yao, "A network analysis of the Chinese stock market," *Physica A, Stat. Mech. Appl.* vol. 388, pp. 2956–2964, Jul. 2009.
[38] T. H. Roh, "Forecasting the volatility of stock price index," *Expert Syst. Appl.*, vol. 33, no. 4, pp. 916–922, Nov. 2007.
[39] M. A. Kaboudan, "Genetic programming prediction of stock prices," *Comput. Econ.*, vol. 16, no. 3, pp. 207–236, 2000.
[40] C. B. Kalayci, O. Ertenlice, H. Akyer, and H. Aygoren, "An artificial bee colony algorithm with feasibility enforcement and infeasibility toleration procedures for cardinality constrained portfolio optimization," *Expert Syst. Appl.*, vol. 85, pp. 61–75, Nov. 2017.
[41] V. Kansal and R. Kumar, "Optimized feature extraction based artificial 1301 intelligence technique for empirical analysis of stock market data," *Int. J. Innov. Technol. Exploring Eng.*, vol. 8, no. 10, Aug. 2019.
[42] Y. Kara, M.A. Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange," *Expert Syst. Appl.* vol. 38, no. 5, pp. 5311–5319, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417410011711
[43] M. Kaucic, M. Moradi, and M. Mirzazadeh, "Portfolio optimization by improved NSGA-II and SPEA 2 based on different risk measures," *Financial Innov.*, vol. 5, no. 1, pp. 1–28, Dec. 2019, doi: 10.1186/s40854-019-0140-6.
[44] M.-J. Kim, S.-H. Min, and I. Han, "An evolutionary approach to the combination of multiple classifiers to predict a stock price index," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 241–247, Aug. 2006.
[45] S. H. Kim and S. H. Chun, "Graded forecasting using an array of bipolar predictions: Application of probabilistic neural networks to a stock market index," *Int. J. Forecasting*, vol. 14, no. 3, pp. 323–337, Sep. 1998.
[46] R. Kizys, A. Juan, B. Sawik, and L. Calvet, "A biased-randomized iterated local search algorithm for rich portfolio optimization," *Appl. Sci.*, vol. 9, no. 17, p. 3509, Aug. 2019.
[47] R. J. Kuo, C. H. Chen, and Y. C. Hwang, "An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network," *Fuzzy Sets Syst.*, vol. 118, no. 1, pp. 21–45, Feb. 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165011498003996
[48] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stock price return via sentiment analysis," *Knowl.-Based Syst.*, vol. 69, no. 1, pp. 14–23, Oct. 2014.
[49] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
[50] A. W. Lo, "Efficient markets hypothesis," in *The New Palgrave Dictionary of Economics*. Basingstoke, U.K.: Palgrave Macmillan, 2007.
[51] W. Long, Z. Lu, and L. Cui, "Deep learning-based feature engineering for stock price movement prediction," *Knowl.-Based Syst.*, vol. 164, pp. 163–173, Jan. 2019.
[52] R. Mansini and M. G. Speranza, "Heuristic algorithms for the portfolio selection problem with minimum transaction lots," *Eur. J. Oper. Res.*, vol. 114, no. 2, pp. 219–233, Apr. 1999.
[53] R. Mansini, W. Ogryczak, and M. G. Speranza, "Twenty years of linear programming based portfolio optimization," *Eur. J. Oper. Res.*, vol. 234, no. 2, pp. 518–535, Apr. 2014.
[54] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1952.tb01525.x
[55] M. A. Mittermayer, "Forecasting intraday stock price trends with text mining techniques," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 37, 2004, pp. 1029–1038.
[56] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," in *Proc. IEEE 5th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Apr. 2019, pp. 205–208.
[57] R. Moral-Escudero, R. Ruiz-Torrubiano, and A. Suárez, "Selection of optimal investment portfolios with cardinality constraints," in *Proc. IEEE Int. Conf. Evol. Comput.*, Jul. 2006, pp. 2382–2388.
[58] J. J. Murphy, *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications* (New York Institute of Finance Series). New York, NY, USA: New York Institute of Finance, 1999. [Online]. Available: https://books.google.com.br/books?id=5zhXEqdr_IcC
[59] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, and E. Salwana, "Deep learning for stock market prediction," *Entropy*, vol. 22, no. 8, p. 840, Jul. 2020.
[60] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Syst. Appl.*, vol. 42, no. 24, pp. 9603–9611, Dec. 2015.
[61] N. Oliveira, P. Cortez, and N. Areal, "Stock market sentiment lexicon acquisition using microblogging data and statistical measures," *Decis. Support Syst.*, vol. 85, pp. 62–73, May 2016.
[62] V. Pagolu, K. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," in *Proc. Int. Conf. Signal Process., Commun., Power Embedded Syst.*, 2017, pp. 1345–1350.
[63] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 259–268, Jan. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417414004473
[64] N. M. Pindoriya, S. N. Singh, and S. K. Singh, "Multi-objective mean–variance–skewness model for generation portfolio allocation in electricity markets," *Electr. Power Syst. Res.*, vol. 80, no. 10, pp. 1314–1321, Oct. 2010.
[65] A. Ponsich, A. L. Jaimes, and C. A. C. Coello, "A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications," *IEEE Trans. Evol. Comput.*, vol. 17, no. 3, pp. 321–344, Jun. 2013.
[66] J.-Y. Potvin, P. Soriano, and M. Vallée, "Generating trading rules on the stock markets with genetic programming," *Comput. Oper. Res.*, vol. 31, no. 7, pp. 1033–1047, Jun. 2004.
[67] R. Ren, D. D. Wu, and T. Liu, "Forecasting stock market movement direction using sentiment analysis and support vector machine," *IEEE Syst. J.*, vol. 13, no. 1, pp. 760–770, Mar. 2019.
[68] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *J. Banking Finance*, vol. 26, no. 7, pp. 1443–1471, Jul. 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378426602002716
[69] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *J. Risk*, vol. 2, no. 3, pp. 21–41, 2000.
[70] F. Z. Xing, E. Cambria, and Y. Zhang, "Sentiment-aware volatility forecasting," *Knowl.-Based Syst.*, vol. 176, pp. 68–76, Jul. 2019.
[71] Y. Ruan, A. Durresi, and L. Alfantoukh, "Using Twitter trust network for stock market analysis," *Knowl.-Based Syst.*, vol. 145, pp. 207–218, Apr. 2018.
[72] K. Schierholt and C. H. Dagli, "Stock market prediction using different neural network classification architectures," in *Proc. IEEE/IAFE Conf. Comput. Intell. Financial Eng.*, Mar. 1996, pp. 72–78.
[73] R. P. Schumaker and H. Chen, "A discrete stock price prediction engine based on financial news," *Computer*, vol. 43, no. 1, pp. 51–56, Jan. 2010.
[74] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The AZFin text system," *ACM Trans. Inf. Syst.*, vol. 27, no. 2, pp. 1–19, Feb. 2009.
[75] S. M. Seyedhosseini, M. J. Esfahani, and M. Ghaffari, "A novel hybrid algorithm based on a harmony search and artificial bee colony for solving a portfolio optimization problem using a mean-semi variance approach," *J. Central South Univ.*, vol. 23, no. 1, pp. 181–188, Jan. 2016.
[76] Y. L. T. V. Silva, A. B. Herthel, and A. Subramanian, "A multi-objective evolutionary algorithm for a class of mean-variance portfolio selection problems," *Expert Syst. Appl.*, vol. 133, pp. 225–241, Nov. 2019.
[77] P. C. Sinha, "Stocks' pricing dynamics and behavioral finance: A review," *Manage. Sci. Lett.*, vol. 5, no. 9, pp. 797–820, 2015.
[78] J. Smailović, M. Grčar, N. Lavrač, and M. Žnidaršič, *Predictive Sentiment Analysis of Tweets: A Stock Market Application* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7947. Berlin, Germany: Springer, 2013, pp. 77–88.
[79] M. G. Speranza, "A heuristic algorithm for a portfolio optimization model applied to the milan stock market," *Comput. Oper. Res.*, vol. 23, no. 5, pp. 433–441, May 1996.

[80] R. E. Steuer and P. Na, "Multiple criteria decision making combined with finance: A categorized bibliographic study," *Eur. J. Oper. Res.*, vol. 150, no. 3, pp. 496–515, Nov. 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221702007749
[81] R. Subbu, P. Bonissone, N. Eklund, S. Bollapragada, and K. Chalermkraivuth, "Multiobjective financial portfolio design: A hybrid evolutionary approach," in *Proc. IEEE Congr. Evol. Comput.*, vol. 2, Sep. 2005, pp. 1722–1729.
[82] W. Sun, C. Tian, and G. Yang, "Network analysis of the stock market," Stanford Univ., Stanford, CA, USA, Tech. Rep. 70, 2015.
[83] S. Takahashi, M. Takahashi, H. Takahashi, and K. Tsuda, *Analysis of Stock Price Return Using Textual Data and Numerical Data Through Text Mining* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 4252. Berlin, Germany: Springer, 2006, pp. 310–316.
[84] S. Takahashi, M. Takahashi, H. Takahashi, and K. Tsuda, *Analysis of the Relation Between Stock Price Returns and Headline News Using Text Categorization* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 4693. Berlin, Germany: Springer, 2007, pp. 1339–1345.
[85] C. K. Vignesh, "Applying machine learning models in stock market prediction," *EPRA Int. J. Res. Develop.*, vol. 5, no. 4, pp. 395–398, Apr. 2020.
[86] B. Wang, H. Huang, and X. Wang, "A novel text mining approach to financial time series forecasting," *Neurocomputing*, vol. 83, pp. 136–145, Apr. 2012.
[87] Y. Wang, "Predicting stock price using fuzzy grey prediction system," *Expert Syst. Appl.*, vol. 22, no. 1, pp. 33–38, Jan. 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417401000471
[88] Y. Wang, "Mining stock price using fuzzy rough set system," *Expert Syst. Appl.*, vol. 24, no. 1, pp. 13–23, Jan. 2003.
[89] B. Wuthrich, V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, and J. Zhang, "Daily stock market forecast from textual Web data," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, vol. 3, Oct. 1998, pp. 2720–2725.
[90] W. Zadrozny, "A comparison of neural network methods for accurate sentiment analysis of stock market tweets," in *Proc. ECML PKDD Workshops, MIDAS PAP*, vol. 11054, Dublin, Ireland. Cham, Switzerland: Springer, Sep. 2019, p. 51.
[91] W.-G. Zhang, Y.-J. Liu, and W.-J. Xu, "A possibilistic mean-semivariance-entropy model for multi-period portfolio selection with transaction costs," *Eur. J. Oper. Res.*, vol. 222, no. 2, pp. 341–349, Oct. 2012.
[92] L. Malandri, F. Z. Xing, C. Orsenigo, C. Vercellis, and E. Cambria, "Public mood–driven asset allocation: The importance of financial sentiment in portfolio management," *Cognit. Comput.*, vol. 10, no. 6, pp. 1167–1176, Dec. 2018.
[93] Z. Zhanggui, H. Yan, and A. M. Fu, "A new stock price prediction method based on pattern classification," in *Proc. Int. Joint Conf. Neural Netw.*, vol. 6, 1999, pp. 3866–3870.
[94] X. Zhong and D. Enke, "Predicting the daily return direction of the stock market using hybrid machine learning algorithms," *Financial Innov.*, vol. 5, no. 1, pp. 1–20, Dec. 2019.
[95] F. Z. Xing, E. Cambria, and R. E. Welsch, "Intelligent asset allocation via market sentiment views," *IEEE Comput. Intell. Mag.*, vol. 13, no. 4, pp. 25–34, Nov. 2018.
[96] H. Zhu, Y. Wang, K. Wang, and C. Y. Chen, "Particle Swarm Optimization (PSO) for the constrained portfolio optimization problem," *Expert Syst. Appl.* vol. 38, no. 8, pp. 10161–10169, 2011.
[97] J. M. T. Wu, Z. Li, G. Srivastava, M. H. Tasi, and J. C.-W. Lin, "A graph-based convolutional neural network stock price prediction with leading indicators," *Softw Pract Exper.*, vol. 2020, pp. 1–17, Oct. 2020.
[98] H. H. Tsai, M. E. Wu, and W. H. Wu, "The information content of implied volatility skew: Evidence on Taiwan stock index options," *Data Sci. Pattern Recognit.*, vol. 1, no. 1, pp. 48–53, 2017.

FERNANDO G. D. C. FERREIRA received the degree in computer engineering and the master's degree in mathematical and computational modeling. He is currently pursuing the Ph.D. degree in mathematical and computational modeling with the Federal Center for Technological Education of Minas Gerais, Brazil. He has been a Visiting Scholar with the University of Technology Sydney, Australia, since January 2020. He has also been doing research in Artificial Intelligence applied to financial investments since 2016.



AMIR H. GANDOMI (Senior Member, IEEE) is currently a Professor of Data Science and an ARC DECRA Fellow with the Faculty of Engineering and Information Technology, University of Technology Sydney. Prior to joining UTS, he was an Assistant Professor at Stevens Institute of Technology, Hoboken, NJ, USA, and a Distinguished Research Fellow in BEACON Center, Michigan State University, East Lansing, MI, USA. He has published over 200 journal papers and seven books which collectively have been cited over 19,000 times (H-index = 64). He has been named as one of the most influential scientific mind and Highly Cited Researcher (top 1% publications and 0.1% researchers) for four consecutive years, from 2017 to 2020. He also ranked 18th in GP bibliography among more than 12,000 researchers. He is active in delivering keynotes and invited talks. His research interests are global optimization and (big) data analytics using machine learning and evolutionary computations in particular. He has served as an Associate Editor, Editor, and Guest Editor in several prestigious journals such as AE of SWEVO, IEEE TBD, and IEEE IoTJ.

RODRIGO T. N. CARDOSO is currently an Associate Professor with the Department of Mathematics, Federal Center for Technological Education of Minas Gerais, Brazil. He has experience in applied mathematics, with emphasis on the following subjects: epidemics control, investment portfolios, multiobjective optimization, and impulsive control.