Expert Systems With Applications 259 (2025) 125299

Contents lists available at ScienceDirect

Expert Systems With Applications

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

An Advisor Neural Network framework using LSTM-based Informative Stock Analysis

Fausto Ricchiuti, Giancarlo Sperlí *

Department of Electrical Engineering and Information Technology (DIETI), University of Naples Federico II, Via Claudio 21, 80125, Naples (NA), Italy

ARTICLE INFO

Keywords:
Stock forecasting
Stock selection
Stock Market
Deep Learning

ABSTRACT

In the past years, the widespread diffusion of Artificial Intelligence (AI) in the finance domain transformed different services, with particular attention to the stock market. Although different AI-based approaches have been proposed for stock forecasting, they are focused on news content or sentiment without considering fundamental features and vice versa. In turn, other approaches rely on handmade rules or ones based on technical indicators for providing advice without considering contextual information that can strongly affect the stock market. In this paper, we propose an Advisor Neural Network framework using Long Short-Term Memory (LSTM)-based Informative Stock Analysis for *Daily investment Advice*. Specifically, the forecasting unit relies on a LSTM-based model, which combines technical indicators, contextual information, and financial data for stock forecasting. Successively, the advice unit provides next-day advice based on predicted information in conjunction with the proposed *Heuristic Stocks Selection* algorithm. This framework has been evaluated on the Stock and Cryptocurrencies markets, considering a subset of 417 stocks and 67 cryptocurrencies over three years, respectively. We compared the proposed framework with several state-of-the-art approaches, showing how it outperforms the baseline in both markets. Furthermore, we achieved a financial gain greater than 41%, despite the downward trend of the NASDAQ market in the quarter under review, and we obtained a 39.38% return on investment for the Cryptocurrencies market.

1. Introduction

The stock market serves as a public marketplace where individuals can buy and/or sell shares of companies at a mutually agreed price (Htun et al., 2023; Thakkar & Chaudhari, 2021a). Hence, it plays a key role in the modern economic system generating higher profits than other financial services although it involves higher risk (Chen et al., 2023; Yi et al., 2023; Zhang et al., 2022). In fact, market capitalization in September 2022 was around $104 trillion. The world's economic growth was anticipated to fall from 5.7% in 2021 to 4.1% in 2022, but it now seems that such estimates were too optimistic, as we are now facing a 2.5% decrease in global growth.[1] Hence, stock market forecasting is one the main important economic tasks since an accurate prediction of stock trends can mitigate the risks for investors, enabling them to make more and more informed decisions about investments.

A first classification of state-of-the-art techniques has been proposed by Henrique et al. (2019), classifying them into two categories: (i) *Technical Analysis* (TA), which uses indicators calculated from the past to indicate future trends of the market; and (ii) *Fundamental Analysis* (FA), which seeks economic factors that influence market trends.

From another point of view, these approaches can be classified into econometric models and machine learning methods (Chen et al., 2021b). The former is composed of stochastic models (e.g., AR Zolfaghari & Gholami, 2021, ARIMA Pai & Lin, 2005, GARCH Wang et al., 2020), whose aim is to identify linear or approximate patterns in stock data (Ding & Li, 2021; Sampaio & Morettin, 2020) although the non-linearity and volatile of the stock market pose several challenges in their applicability (Jin et al., 2022; Paiva et al., 2019; Wu et al., 2022). Despite some empirical studies (Cervelló-Royo & Guijarro, 2020; Henrique et al., 2019) have supposed that the financial market can be predictable to some extent, contextual information strongly exerts information on the stock markets.

In turn, data-driven methodologies, mainly based on machine learning (Kumbure et al., 2022), have been proposed to deal with the non-linearity, noise, and non-stationarity of stock trend (Yang et al., 2021). A literature review has been discussed in Nazareth and Ramana Reddy (2023) for providing an overview of machine learning models for several financial applications. In turn (Olorunnimbe & Viktor, 2022) investigated deep learning models for the stock market, also focusing

* Corresponding author.
E-mail addresses: f.ricchiuti@studenti.unina.it (F. Ricchiuti), giancarlo.sperli@unina.it (G. Sperlí).

[1] https://fortunly.com/statistics/stock-market-statistics/#gref

https://doi.org/10.1016/j.eswa.2024.125299
Received 4 January 2024; Received in revised form 21 August 2024; Accepted 1 September 2024
Available online 3 September 2024

Table 1
Nomenclature.

Acronym	Description
Bi-LSTM	Bidirectional Long Short-Term Memory
CNN	Convolution Neural Network
CSUM	Cumulative Sum
DR	Daily Return
FC	Fully Connected
FL	Fuzzy Logic
GCNN	Graph Convolutional Neural Network
GRU	Gated Recurrent Unit
G_i	Cumulative Percentage Economical Gain at day i
K	Number of suggested stocks
LDA	Latent Dirichlet Allocation
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Average Error
MSE	Mean Squared Error
ML	Machine Learning
NN	Neural Network
ReLu	Rectified Linear Unit
RMSE	Root Mean Squared Error
S	Set of Stocks
SGD	Stochastic Gradient Descent
STKD LSTM	Stacked Long Short-Term Memory
STKD BI-LSTM	Stacked Bidirectional Long Short-Term Memory
STKD GRU	Stacked Gated Recurrent Unit
STKD BI-GRU	Stacked Bidirectional Gated Recurrent Unit
SVM	Support Vector Machine
X_i	Capital at day i

on the use of historical data for retrospective evaluation of models' effectiveness (*backtesting*).

Stock forecasting has attracted increasing attention from researchers who have developed approaches (Nazareth & Ramana Reddy, 2023; Tang et al., 2022) although different challenges are still open (Ge et al., 2022; Thakkar & Chaudhari, 2021a).

Model- and statistical-based approaches were the first ones used for stock forecasting although they encountered different challenges in being applied to the financial domain. The former encounters difficulty in modeling the intrinsic relationships in financial data while the latter struggles to capture non-linear and non-stationary behavior of the stock market (de A. Araújo & Ferreira, 2013; Wang et al., 2022a).

Although Artificial Intelligence models have been designed to deal with these issues, their effectiveness is affected by internal and external factors. Possible internal challenges mainly concern the feature selection process to support forecasting model (Htun et al., 2023) and the choice of a suitable evaluation measure (Olorunnimbe & Viktor, 2022). Furthermore, the complexity of the stock forecasting task is strongly increased by the analysis of a large amount of multimedia content published on Online Social Networks and news to infer user's opinion and/or feelings (Chen et al., 2022; Choi et al., 2024; Jing et al., 2021). Another challenge concerns how the unforeseeable events may affect the effectiveness of stock forecasting approaches (Dong et al., 2022; Henrique et al., 2023).

Hence, external factors (i.e., people's feelings, experts' opinions and security breaches) (Ashtiani & Raahemi, 2023; Thakkar & Chaudhari, 2021b), as well as the intrinsic non-linear and non-stationary behavior of the stock market (Agarwal et al., 2019; Htun et al., 2023), pose several challenges in defining suitable models. In addition, financial data heterogeneity requires more and more sophisticated methodologies based on deep learning to deal with the non-linear and non-stationary market trends (Kehinde et al., 2023; Park & Yang, 2024).

In this paper, we propose an Advisor Neural Network framework using Long Short-Term Memory ($LSTM$)-based Informative Stock Analysis for producing *Daily investment Advice*. It is mainly composed of two phases: we first forecast the stock market through a $LSTM$-based model that combines technical indicators, contextual information, and financial data while providing next-day suggestions. We further integrate seasonal data into the proposed model to learn the seasonality behavior of the stock market. Summarizing, technical indicators and contextual information aim to summarize the behavior of stock modeled as a time series and to jointly investigate the burst of financial news and the related daily average sentiment of each stock, respectively. In turn, historical financial information provides information about the daily price and volume of each stock while seasonal data aims to learn the seasonality behavior of the stock market. This information is combined as a time series into the $LSTM$-based model to predict the next-day price. Successively, we propose an innovative *Heuristic Stocks Selection* algorithm based on the deep learning module output to handle the risk of losing money during the investment procedure. Its magnitude of the positive score is inversely proportional to the absolute prediction error while the negative score increases with the forecast error.

The framework has been evaluated on two datasets, composed of more than 400 stocks in the NASDAQ and 67 Cryptocurrencies collected over a three-year period, to provide investment suggestions for each market day in a trimester (August, September, and October 2022) by investing the capital divided equally on each suggested stock. It has been evaluated in terms of daily return to consider the risk of investing money in the stock market. It is worth noting that we buy and sell stocks at the market's opening and closing respectively, without considering trading fees. As an outcome of the experimental evaluation, we achieved an increase of investment capital greater than 43%, despite the downward trend of the NASDAQ market in the quarter under review. Furthermore, the same simulation has been replicated on a three-year dataset of 67 Cryptocurrencies, obtaining a 39.38% return on initial capital. In Table 1, we summarize the basic notation used in this paper.

Summarizing, the main novelties of the proposed framework concerns:

- A novel *Deep Learning Framework* for Stock Market Forecasting, which combines contextual information extracted from news with historical financial data in conjunction with technical analysis, whose selection may improve the effectiveness of the stock price movements, as underlined in Ji et al. (2022) and Thakkar and Chaudhari (2021b). We perform a comprehensive feature selection process to identify the most relevant set of features to support the forecast model. In particular, we are interested in investigating news content because they are most relevant for stock forecasting as underlined in Dong et al. (2022);
- An innovative *Heuristic Stocks Selection Algorithm* that significantly increases advisory performance how to measure the instability of stock data and stock selection are not considered (Bai et al., 2023). The advice strategy relies on this selection algorithm, whose magnitude of the positive score is inversely proportional to the absolute prediction error while the negative score increases with the forecast error. Hence, it allows us to handle the uncertainty in forecasting financial markets, whose behaviors are uncertainty and volatility;
- An experimental evaluation on two real-world scenarios composed by three-year datasets about stock and cryptocurrency markets for providing daily investment advice: the former composed of 400 NASDAQ market shares and the latter made up of 67 Cryptocurrencies. We further compare the proposed framework with several state-of-the-art approaches, on both markets.

The remainder of the paper is organized into six sections. Section 2 investigates the State-of-the-Art approaches for stock market forecasting while Section 3 describes the proposed framework, which combines historical financial data and contextual information and can be used as an investment advisor. Section 4 discusses the experimental analysis made on two different real-world scenarios (stock market and cryptocurrencies), that collects three years of information, to evaluate the proposed framework's effectiveness. Finally Section 5 summarizes the main findings of the proposed analysis, also identifying possible directions for future works.

2. Related works

The analysis of stock markets has gathered significant interest from both practitioners and researchers (Kehinde et al., 2023; Li et al., 2018) in order to investigate the dynamic nature of stock prices to jointly mitigate investment risks and enhance profitability. In a recent study, Kehinde et al. (2023) performed a scientometric review, analyzing 220 articles published between 2001 and 2021 to identify trends and patterns in stock market forecasting research, also unveiling the relevance of machine-learning keywords in the context of stock market analysis.

Nevertheless, the intrinsic non-linear and non-stationary behavior of the stock market poses challenges to the prediction of prices or trends (de A. Araújo & Ferreira, 2013; Wang et al., 2022a). To deal with these issues, researchers have focused their efforts on designing several approaches, that can be classified into three categories: (i) model-based (Cheng & Yang, 2018; Wei, 2016), (ii) statistical-based (Efendi et al., 2018; Hafiz et al., 2023) and (iii) data-driven methods (Kumbure et al., 2022). The challenges of the former encounter in formulating a model capable of capturing the intricate relationships within financial data. Conversely, the latter confronts the issues arising from the non-linear and non-stationary nature of financial data, despite its apparent advantages in computational efficiency and accuracy.

In turn, the third category has attracted increasing interest since Artificial Intelligence (AI) models have been applied to increasingly broader areas of finance, whose opportunities and challenges have been discussed by Cao (2022). Specifically, stock forecasting has become a major investigated financial area through the machine and deep learning models (see Nazareth & Ramana Reddy, 2023 for a comprehensive analysis). Furthermore, Nazareth and Ramana Reddy (2023) examine data characteristics and selected features for evaluating these models while (Olorunnimbe & Viktor, 2022) reviewed deep learning-based approaches involving backtesting. In Thakkar and Chaudhari (2021a), the authors surveyed Deep Neural Network (DNN) models to deal with stock forecasting, also investigating hybrid and metaheuristic approaches with DNNs. Ashtiani and Raahemi (2023) have discussed a further literature review on stock market prediction through AI-based approaches by considering both numerical (i.e., stock and technical indicator) and textual (i.e., news and sentiment) data.

Numerous studies have proved that $LSTM$ outperforms most single econometric machine learning, and neural network models (Kim & Won, 2018) when it comes to stock price forecasting (Ahmed et al., 2022; Zhong et al., 2023). This observation underlines the inherent capacity of $LSTM$ models to handle time series forecasting task effectively.

However, these AI-based models pose several challenges to their design and deployment in real-world scenarios. The first challenge concerns the feature selection process to support the forecasting model, as underlined by Htun et al. (2023). Despite the majority of approaches relying only on the use of technical indicators (Chen et al., 2021a) and/or stock features (e.g., volume, open and close prices) (Xie et al., 2021) or their combination (Haq et al., 2021), few studies aim at integrating features related to fundamentals into their analysis (Ozbayoglu et al., 2020). Other approaches (Huang et al., 2023) focus the analysis on the possible correlations between companies through a graph-based model by combining historical financial data with industry-centered information (i.e., Location of a company's headquarters, product similarity, assets, and dividend). Nevertheless, these approaches do not consider contextual features (i.e., social networks and/or news content) that are suitable to handle unforeseen events that might affect effectiveness performances.

Another challenge concerns the choice of evaluation metrics, as outlined by Dessain (2022), in which the authors stated that risk-adjusted return measures may be preferable over others (e.g., Mean Squared Error (MSE), Root Mean Squared Error (RMSE)) although the former suffers from statistical limitations. In turn, Olorunnimbe and Viktor (2022) recognize that domain-specific metrics such as "returns" and "volatility" appear most important for accurately representing model performance across specializations.

Although AI models are developing rapidly (Huang et al., 2023; Ma et al., 2022), unforeseen events limit their predictive power as shown in Ahelegbey et al. (2022), Gangopadhyay and Majumder (2023), Gjerstad et al. (2021) and Ronaghi et al. (2022). Hence, researchers and practitioners have strived to predict the financial market by analyzing textual (e.g., news articles and social media) and numeric data (e.g., hourly stock prices, and moving averages) (Li et al., 2020; Picasso et al., 2019; Schmitz et al., 2023). However, Dong et al. (2022) investigated the predictive value of news and social media finding that the former is most effective with a one-day horizon, while the latter can support predictions over two to five days.

Despite different deep learning-based approaches (Anbaee Farimani et al., 2022; Chen et al., 2019; Maqsood et al., 2020) have been proposed, they mainly focus on analyzing news content or sentiment without investigating features associated with fundamentals. In turn, other methodologies aim to integrate investment scores based on technical indicators (Banik et al., 2022) or infer rules based on historical financial data (Ozcalici & Bumin, 2022) without considering contextual information that can strongly affect the forecasting task.

Although several state-of-the-art approaches (see Table 2 for their main characteristics) have been proposed to deal with the stock forecasting task, they suffer from some drawbacks that the proposed approach aims to overcome. The majority of the state-of-the-art approaches (see for instance Anbaee Farimani et al., 2022) mainly relies on manually selected news from one of domain specific portals, limiting the number of news records significantly impairs the performance of machine learning models, as shown in Ashtiani and Raahemi (2023). Other approaches (see for instance (Banik et al., 2022)) aims to predict stock behavior over 30 days without considering contextual information (i.e., news information) that affect the stock behavior. In turn, other methodologies (see for instance Gangopadhyay & Majumder, 2023) focus solely on predicting the movement of closing prices by integrating news content embeddings with other financial information, without assessing the impact of news on the future value of the stocks. Furthermore, other approaches (see for instance (Bai et al., 2023)) do not consider stock selection algorithm that can increase advisory performances taking into account uncertainty in stock behavior.

Our proposal relies on the design of an innovative advice strategy integrating prediction of deep learning models. This model aims at performing stock forecasting by combining historical financial information, technical indicators, seasonal data, and contextual information, whose combination has not been widely explored, as shown in Table 2, into a forecasting model to support an advice strategy based on an innovative Heuristic Stocks Selection algorithms. We further enhance the proposed model by integrating both the daily number of news articles and the daily average sentiment score for each stock through the collection of a large amount of news data. Once collected all the information, they are modeled as a multivariate time series to be fed as input to LSTM-based methodology. Successively, the advice strategy relies on a novel stock selection algorithm, whose magnitude of the positive score is inversely proportional to the absolute prediction error while the negative score increases with the forecast error.

Summarizing:

- The proposed framework aims to combine historical financial information with news content and seasonal data w.r.t. Anbaee Farimani et al. (2022), Banik et al. (2022), Chen et al. (2019), Maqsood et al. (2020), that are only focused on news content or sentiment score respectively, without analyzing fundamental features;
- The proposed framework relies on the proposed heuristic selection algorithm on the basis of the deep learning module output, to identify stocks to invest in w.r.t. approaches based on handmade rules (Ozcalici & Bumin, 2022) or technical indicators (Banik et al., 2022; Huang et al., 2023);

Table 2
Analysis of the state-of-the-art approaches on the basis of the used features (i.e., Historical Data (HD), News (Nw), Social Media (SM), Technical Indicator (TI), Seasonality (Se)), the related prediction model and the predicted variables.

Approaches	Features					Predictors	Predicted Variable
	HD	Nw	SM	TI	Se		
Li et al. (2018)	Close Price	News Count, Sentiment Score				ML, Regression and Statistical Model	Volatility Index
Wang et al. (2022a)	Close Price					LSTM	Closing Price
Cheng and Yang (2018)	Close Price					Fuzzy Logic	Closing Price
Wei (2016)	Close Price					Fuzzy Inference	Closing Price
Efendi et al. (2018)	Low-High Price					Fuzzy random auto regression model	Closing Price
Ghosh et al. (2022)	Close Price					LSTM and Random Forest	Stock Price
Hafiz et al. (2023)	Close Price			EMA, RMSI, MACD, LWO, PO, DIU, DID, A ratio, B ratio, ROC, UO, Ulcer, MTM		Co-evolution of neural architectures and features	Closing Price
Thakkar and Chaudhari (2021a)	Close Price					DNN	Adjusted Close Price
Ashtiani and Raahemi (2023)		Sentiment Score				ML	Closing Price
Kim and Won (2018)	Close price					LSTM + GARCH	Volatility index
Ahmed et al. (2022)	Close price					Regression and Stacked layers of LSTM	Closing Price
Zhong et al. (2023)	Trading volume, market capitalization, OHLC[a] price data.			Last five lagged returns, EMA, CSUM of three and five days, RSI, William's percentage, MACD, OBV.		LSTM + relationwise graph attention network	Stock binary label
Chen et al. (2021a)	Close price			1-day ROC, 5-day TEMA, 5-day WMA, 5-day SMA, 5-day HMA, 5-day EMA, 9-day CMO, 9-day WR, MACD histogram and 14-day CCI		GCNN	Stock binary label
Xie et al. (2021)	Close price					Neuro-fuzzy System	Closing Price
Haq et al. (2021)	OHLC price data Median, Mean and Typical prices			EMA, MA, KAMA, MACD, MACDH, PLUS-DI, PLUS-DM,CCI, PPO,DX, ADX,ADXR,ATR,NATR		Generative Model with Attention	Stock binary label
Huang et al. (2023)	Close price					Gated Graph Neural Network with attention	Stock binary label
Ronaghi et al. (2022)	Close price		Tweet Embedding			CNN + Bi-LSTM	Adjusted Close Price
Gangopadhyay and Majumder (2023)	OHLC price data	News Embedding		ADX, MACD, MOM, ATR, RSI, STOCH, WILLR, BBANDS, EMA, SMA		FC Layer with Softmax Activation	Stock binary label
Gjerstad et al. (2021)			Bigrams			LDA	Return in a window
Li et al. (2020)	Close price	Sentiment score		MA for 10,20,30 days DIFF, DEA, MACD, RSI for 6,12,24 days and MFI		LSTM	Price return rate
Picasso et al. (2019)	Close price					NN	Stock binary label
Schmitz et al. (2023)	Close price	TF-IDF, Sentiment score				ML	Stock binary label
Chen et al. (2019)	Stock price	Sentiment score				LSTM	Stock binary label
Anbaee Farimani et al. (2022)		Sentiment score		EMA, BB, OBV, ADI,ATR,MACD, AO,RSI		LSTM	Closing Price
Maqsood et al. (2020)	Close price		Sentiment Score			CNN, SVM, LR	Closing Price
Banik et al. (2022)	Closing Price			MACD, RSI,MFI, Support, Resistance		LSTM	Closing Price
Ozcalici and Bumin (2022)	Closing price			Hand-crafted Performance measure		Genetic Algorithm	Adjusted Return
Our proposal	OHLC price data	News Count, Sentiment Score		AO,RSI,ATR, ADX,AI,DR	Day of week and month Index of Month and quarter	LSTM	Daily Return

[a] OHLC state for Open High Low Close prices.

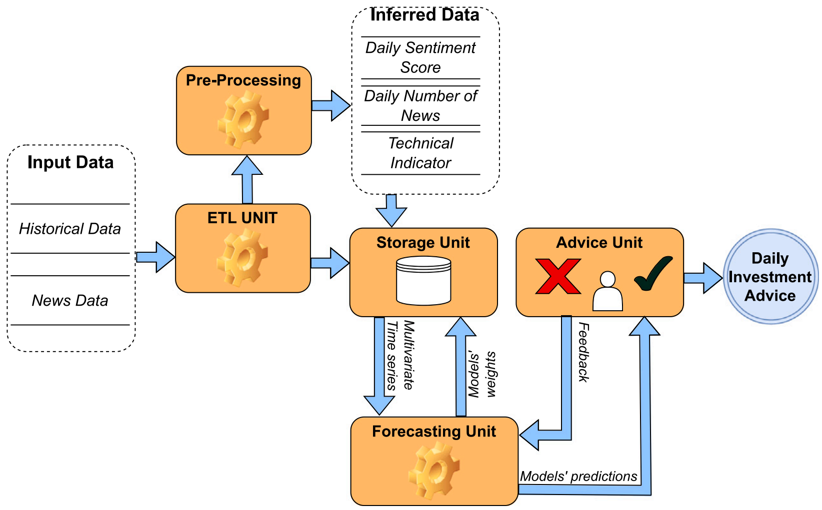

Fig. 1. Overview of the proposed framework, representing the overall flow along with the features (crawled and inferred). It is mainly composed of three modules: (i) *Data Ingestion*, which crawl historical financial data and contextual information that are stored into the storage layer; (ii) *Stock forecasting*, whose aim is to forecast stock trend; and (iii) *Advice suggestions*, which provides a ranking of stocks on the basis of a novel index.

- The proposed framework relies on the $LSTM$-based forecasting module, which has been improved by considering historical financial data with contextual information and technical indicators w.r.t. Kim and Won (2018) which only use historical information. Furthermore, the proposed framework has been statistically evaluated in real-world scenarios in terms of RMSE and MSE, also considering the daily returns that represent a more suitable measure to evaluate the model across the specialization, as shown in Olorunnimbe and Viktor (2022).
- The proposed framework aims to combine historical financial data with contextual news information to handle unforeseen events w.r.t. other approaches (Huang et al., 2023), mainly combining historical financial data with industry-centered information.

3. Framework

The proposed framework, shown in Fig. 1, aims to provide daily stock market advice, which can be used by investors to improve their financial capital. It is mainly composed of three modules, starting from crawling data until suggesting suggesting a set of stocks: *Data Ingestion*, *Stock Forecasting*, *Advice Suggestion*.

In the first stage, financial data and contextual information are extracted from several heterogeneous sources through Application Programming Interface (API) from different portals, corresponding to the first activity into the flowchart represented in Fig. 2. The crawled information can be classified into two groups: (i) *historical financial data*, mainly related to price and volume, and (ii) *news data*, composed of the daily number of financial news and the related average sentiment for each stock. We, further, infer *Daily Sentiment Score* and *Daily number of news* from textual content data as well as *technical indicator* and *seasonal data* from historical financial data, representing indicators to summarize the trends in the time series and information to represent the possible seasonality of stock, respectively. This stage is represented in the second activity of Fig. 2. These features are, successively, preprocessed to fed them as input in the forecasting layer before storing them in the storage layer (see Section 3.2 for more details). The *forecasting unit*, described in Section 3.3, relies on the $LSTM$ models, that have been trained on the crawled information, stored into the storage layer. Specifically, this information is modeled as a time series to be fed as input to the $LSTM$-based models for stock forecasting, as represented in the fourth activity of Fig. 2. Once the training process is completed, the model weights are saved within the storage layer. Finally, the *Advice Unit* (see Section 3.4) relies on the output of the *Forecasting layer* to provide Daily Investment Advice by ranking the stocks according to a novel index.

3.1. Task

In the last years, researchers and practitioners have focused on developing several models and approaches for Stock Market Forecasting. This focus stems from the significant potential for wealth generation through investment returns that can be harnessed in this area, enabling informed investment decisions. Accurate prediction of stock prices can support investment institutions and investors in attaining substantial profits (Wang et al., 2022a). Nevertheless, the stock market is a challenging domain due to the highly noisy dynamic, non-linear, non-parametric, and chaotic nature of stock data, which can be affected by several factors (i.e., political, macroeconomic, and investment psychology) (Azarnejad & Khaloozadeh, 2022).

Despite different approaches have been proposed in the literature, Dezhkam and Manzuri (2023) classify them into two macro-categories according to the addressed task: *Regression* (see for instance Ahmed et al., 2022) and *classification* (as made in Cagliero et al., 2020). However, the majority of these approaches deal with the first task, being a relevant challenge for professionals and organizations, whose aim is to minimize the forecasting error to make profits in the stock market (Chen et al., 2021a).

Definition 3.1. Stock Forecasting Task Let $S = \{s_1, \ldots, s_N\}$ be a set of N stocks with the related features $\mathcal{T}(s_i) = \{T_1(s_i), \ldots, T_N(s_i)\}$ represented as a set of features, the advisor task concerns the suggestion of a suitable set of stocks, whose next day value (p_{d+1}) is defined according to a regression task ($f : T_1(s_i) \times T_1(s_i) \times \cdots \times T_N(s_i) \rightarrow p_{d+1}$).

Summarizing, stock forecasting (see Definition 3.1) is a complex task in financial analysis, whose aim is to predict the future stock price as a continuous numerical value (Ozbayoglu et al., 2020; Thakkar & Chaudhari, 2021a). Improving stock forecasting aims both to mitigate financial risks for government financial institutions and to reduce losses in the financial market (Kamara et al., 2022; Wang et al., 2021).

3.2. Data ingestion

In this section, we describe the different stages in the *Data Ingestion* module, starting from the crawling phase until the stored information

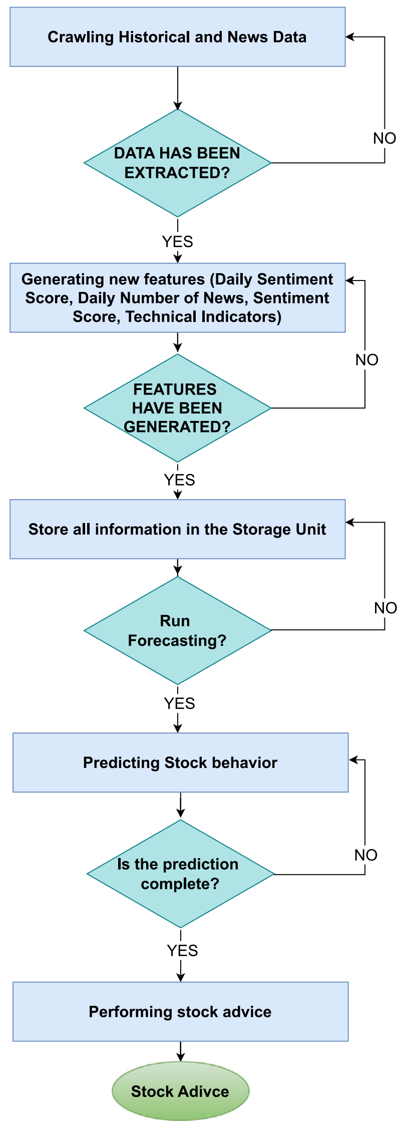

Fig. 2. Flow Diagram of the proposed approach representing its main activities. In the first stage, historical and financial data are crawled from different portals. Once the collection is completed, novel features (i.e., Daily Sentiment score, Daily Number of News, Technical Indicators) are generated. All the information is, then, stored in the Storage Unit, and they are, successively, modeled as a multivariate time series to be fed as input to the LSTM-based models for stock forecasting. Finally, the output of the Forecasting module is integrated into the stock selection algorithms to provide stock advice.

in the *Storage Unit*. Although financial information has been classified by Chen et al. (2021a) into two categories (financial data and technical indicators), we improve the proposed analysis by incorporating an extensive set of contextual information: *Financial Data*, *Technical Indicators*, *Seasonal Data*, *News Data*.

Historical data. In the first stage, we crawl historical finance data from specific portals (e.g., Yahoo Finance,[2] Google Finance[3]), whose main features are summarized in Table 3.

Seasonal data. The second phase concerns the computation of the *Seasonal Data* from the *Date* field of the crawled historical financial market (see Table 3): *Day of the Week*, *Day of the Month*, *Month*, *Quarter*. These categorical features allow the Neural Network to learn the seasonality behavior of the stock market. Once crawled seasonal data, we use the *One-Hot Encoding* (Hancock & Khoshgoftaar, 2020) technique for transforming categorical data into numerical ones.

Table 3
Features extracted from historical market data.

Feature	Meaning
Date	The Market Day to which the row data refer
Open Price	The first price at which buyers and sellers agree to conclude a transaction on a Market Day
High Price	The highest price at which a security is traded during a Market Day
Low Price	The lowest price at which a security is traded during a Market Day
Close Price	The last price at which buyers and sellers agree to conclude a transaction on a Market Day
Volume	Total number of shares traded during the Market Day.

News data. . They are obtained from the *EOD Historical Data*[4] platform, which provides a filtered list of news articles along with associated sentiment scores for each stock. We, further, modeled these data as a multivariate time series, encompassing both the daily count of financial news and the daily average sentiment for each stock. Specifically, the sentiment values are normalized on a scale from $+1$ (completely positive) to -1 (completely negative).

Technical indicator. They are the most commonly used for stock trend prediction, as they effectively summarize behavior or trends within the time series. In the proposed approach, the dataset has been augmented by integrating six technical indicators. Once these technical indicators have been computed, the historical market data have been removed to reduce correlations between features.

The *Awesome Oscillator* (*AO*) measures the momentum of a financial stock (Williams, 1998). Specifically, it is defined as the difference between *Simple Moving Average* (SMA) scores over two different time periods, producing function oscillating around the zero. When the value is above the zero line, it indicates that the short-term moving average is greater than the long-term moving average (*bullish momentum*), and conversely, when below the zero line, it indicates that the short-term moving average is less than the long-term moving average (*bearish momentum*).

The *Relative Strength Index* (*RSI*) is a further technical indicator (in the range [0,100]) which measures the strength of the stock price action (Wilder, Jr., 1986). The RSI is defined as the ratio between upward and downward price movements over a time frame in the form of an oscillator. This indicator is computed through Eq. (1), in which the *Average Gain* (AG) and *Average Loss* (AL) are the sum of the gains and losses over the time period, respectively.

$$RSI = 100 - \frac{100}{1 + \frac{AG}{AL}} \tag{1}$$

A stock is considered *overbought* or *oversold* when the RSI is above 70 or below 30, respectively.

The *Average True Range* (*ATR*) measures the volatility of a financial stock (Wilder, 1978), computed as the average difference between ATR at the previous time instance and the True Range (TR) over a given period of time (n) (see Eq. (2)). High ATR values indicate high levels of stock volatility and vice versa.

$$\frac{ATR(n-1) + TR}{n} \tag{2}$$

[2] https://www.yahoo.com/author/yahoo-finance
[3] https://www.google.com/finance/
[4] https://eodhistoricaldata.com/

Table 4
Overview of $LSTM$ Architecture.

Layer (Type)	Output Shape	Param #
LSTM (LSTM)	(None, 128)	99,328
dropout (Dropout)	(None, 128)	0
dense (Dense)	(None, 128)	16,512
dense_1 (Dense)	(None, 1)	129
Total params: 115,969		
Trainable params: 115,969		
Non-trainable params: 0		

The *Average Directional Movement Index* (ADX) is a Technical Indicator (in the range [0,100]) which measures the strength of a trend in the stock price (Wilder, 1978). It is computed on the basis of the *Positive Directional Indicator* ($+DI$) and *Negative Directional Indicator* (-DI), comparing the current High and Low values with their previous ones and vice-versa. A high ADX value indicates a strong trend.

The *Aroon Indicator* (AI) measures the strength and direction of a trend in the stock's price action (Chande, 2001). It consists of the difference between the *Aroon Up* and the *Aroon Down*, corresponding to the number of periods elapsed since that maximum and minimum are determined, respectively. Furthermore, the former measures the strength of a bullish trend while the latter computes the strength of a bearish trend.

The *Daily Return* (DR) index is the deviation of the stock price on the current day (in percentage), as defined in Eq. (3).

$$DR_i = 100 \cdot \frac{Close_i - Open_i}{Open_i} \quad for\ each\ entry\ i. \tag{3}$$

3.3. Forecasting unit

The *Forecasting Unit* is responsible for designing and updating the Neural Networks utilized in the forecasting process. At this stage, we leverage $LSTM$ models, outperforming to other methods, as underlined by Benidis et al. (2022). One of its main parameters concerns the length of time window, which represents a set of n stock's historical values to be fed as input of $LSTM$ models.

As shown in Fig. 3, this unit is responsible of predicting stocks behavior by designing a model for each stock, whose weights are update on the basis of the novel metric. Hence, this unit consists of three sub-units: (i) *Builder*, (ii) *Update*, and (iii) *Prediction*.

Builder unit. It builds a Neural Network for each stock, represented as a time series, with the aim of predicting its trend. In particular, this module relies on $LSTM$, whose architecture, as shown in Table 4, consists of the following layers:

1. *Input Layer*, which is the entry point of the Neural Network;
2. *LSTM Layer*, which iterates over the Time Series, maintaining an internal state that encodes the information already learned;
3. *Dropout Layer*, which helps prevent overfitting during training time;
4. *Dense Layer*, which is a regular densely-connected Neural Network layer.

We employ different activation functions to train $LSTM$-based networks according to the layer type. We choose *tanh* as the activation function for the $LSTM$ layer. It is a non-linear function that maps its input to the range $[-1, 1]$ (Hochreiter & Schmidhuber, 1997), whose formal definition is shown in Eq. (4). In particular, the *tanh* activation function handles non-linearity in the $LSTM$ network, enabling it to model more complex relationships between the inputs and outputs. It also allows to stabilize the gradients during training, which can improve the performance and speed of training.

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{4}$$

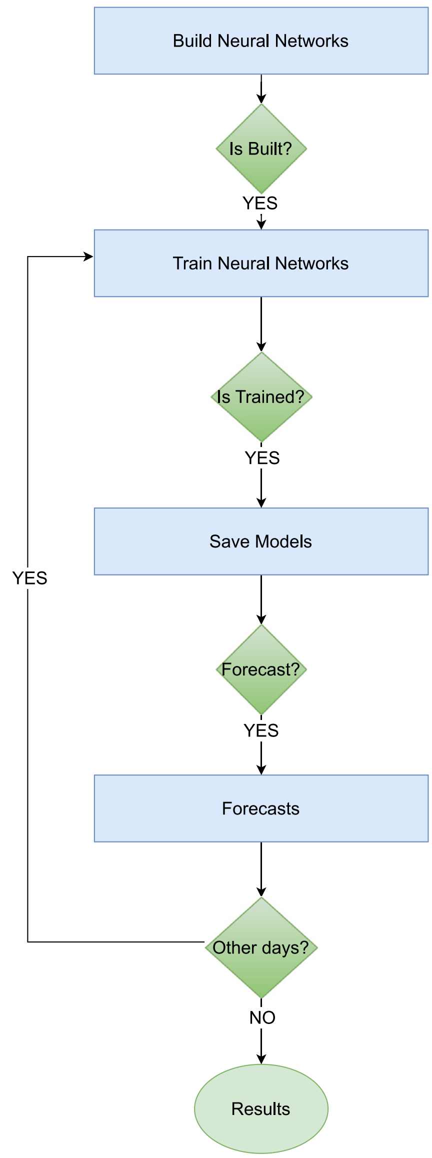

Fig. 3. Flowchart representing the Forecasting Unit, in which we have three different activities: (i) *Design*, whose aim is to design a model for each stock, (ii) *Update*, updating models on the basis of the novel metric, and (iii) *Prediction*, which provides stock forecast.

In turn, the *Rectified Linear Unit* ($ReLU$), defined in Eq. (5), is a commonly chosen activation function for fully connected layers (Nair & Hinton, 2010). It might prevent overfitting by handling sparse activation for large input.

$$ReLU(x) = \max(0, x) \tag{5}$$

The *Linear Activation Function* is a simple function $f(x) = x$, which is commonly used in fully connected layers of Neural Networks (Glorot & Bengio, 2010). It has been chosen as the activation function for the last layer in the proposed model since it is most used in regression problems.

Furthermore, the training phase aims to minimize the *loss* function, defined as the difference between the predicted output and the real value for a given input (Goodfellow et al., 2016). The *Mean Squared Error* has been chosen as the loss function in the developed framework. To pursue this aim, we used the *Optimizer* to find the best set of parameters (weights and bias). In the proposed architecture, we choose the Adaptive Moment Estimation (*Adam Optimizer*) as an optimizer, which is a modified version of *Stochastic Gradient Descent* (SGD) by incorporating adaptive learning and momentum computation (Kingma & Ba, 2014). Once the training process is completed, the model has been saved in the Storage Unit.

Update unit. In the *Update Unit*, a multivariate time series for each stock has been fed as input to a $LSTM$-based Neural Network, which has been evaluated through a novel metric according to Definition (6).

Definition 3.2. Let x and r be respectively the predicted and real values, the score metric has been defined according to Eq. (6).

$$Score = \begin{cases} min(\frac{1}{|x-r|}, MAX) & \text{if } x \cdot r \geq 0 \\ -100 \cdot |x-r| & \text{if } x \cdot r < 0 \end{cases} \tag{6}$$

The designed metric assigns a positive score if the model correctly predicts the positive/negative stock trend, negative otherwise. The magnitude of the positive score is inversely proportional to the absolute prediction error while the negative score increases with the forecast error. Finally, the updated model is loaded into the *Storage Unit* while the model score is sent to the *Advice Unit*.

Prediction unit. The *Prediction Unit* performs the stock value forecast based on the multivariate series fed as input. The predicted score is sent to the *Advice Unit*.

3.4. Advice unit

The *Advice Unit* relies on the prediction score of the *Forecasting Unit* to suggest profitable stocks in which to invest in. Let S be a set of securities, this unit suggests K stocks whose *Daily Return* is expected to be higher on the following day. Hence, the *Advice Unit* is composed of two main components: (i) *Stocks Selection* and (ii) *Advice Production*.

Stocks selection. In this phase, N models are evaluated and updated on multivariate time series of examined stocks through the *Forecasting Unit*. The predicted values are sorted in descending order by choosing the first K ones. Hence, the most trustworthy models on the current day's data are identified, which can be used for the next day's forecasts.

Advice production. The M models selected in the previous step are used to predict the next day's Daily Return by using the corresponding *Forecasting Unit*. The predicted values are sorted in descending order and the first K stocks are selected for the Daily Investment Advice.

4. Experimental analysis

In this section, we describe the experimental analysis made for evaluating the proposed framework according to the experimental protocol (described in Section 4.1) using the evaluation metrics defined in Section 4.2.

4.1. Experimental protocol

The experimental analysis aims to evaluate the effectiveness of the proposed framework in two different scenarios: the stock and the cryptocurrency markets. In particular, the experimental evaluation has a threefold objective:

- Identifying relevant features (i.e., technical indicators, seasonal characteristics, and news data) for stock forecasting (see Section 4.3.1);
- Optimizing framework's parameter according to time window size and neural network type (see Section 4.3.2);
- Evaluating the effectiveness of the proposed framework on 497 stocks and 67 Cryptocurrencies during the 3 months of the test simulation (see Section 4.3.3).

The first two objectives rely on data collected from *01/11/2019* to *31/07/2022* of 40 shares randomly selected from the 500 most capitalized stocks on the NASDAQ Market.[5]

In turn, the last objective is based on a subset of 417 stocks, corresponding to the ones having historical information on NASDAQ before October 2019. For each security, we collected data from *01/11/2019* to *31/10/2022* on which a $LSTM$-based model has been designed for each stock.

Summarizing, the dataset consists of 690 entries for each stock, divided as follows: (i) *Training Set*: 480 entries (from *01/11/2019* to *28/09/2021*), (ii) *Validation Set*: 140 entries (from *29/09/2021* to *19/04/2022*), (iii) *Test Set*: 70 entries (from *20/04/2022* to *29/07/2022*). A similar analysis has been made for the Cryptocurrencies market on a set of 67 Cryptocurrencies over the same period. We summarize the dataset information for stock forecasting in Table 5:

Furthermore, we evaluate advice strategy using a test set spanning from August 1, 2022 to October 31, 2022 while training machine learning models through data collected from November 1, 2019 to July 31, 2022, as shown in Table 6.

The experimental analysis has been carried out on Google Colab,[6] a Platform-as-a-Service (PaaS) made up of an Intel Xeon CPU @2.30 GHz and 13 GB RAM, by using a technology stack based on Python 3.9 with different machine learning and deep learning libraries (e.g., *Sklearn*,[7] *Tensorflow*,[8] and *Keras*[9]).

4.2. Evaluation metrics

To evaluate the performance of the predictive model, we compare the real value (y_i) with the predicted one ($\hat{y}_i$) for the ith observation, whose total number is n, in order to measure the error made by the model.

The *Mean Absolute Error* (*MAE*) is defined as the average of the absolute differences between the predicted and actual values of a variable (Abdi, 2007), as shown in Eq. (7).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{7}$$

The *Mean Squared Error* (*MSE*) is measured as the average of the squared differences between the predicted and actual values of a variable (Bishop & Nasrabadi, 2006), as shown in Eq. (8).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{8}$$

The *Root Mean Squared Error* (*RMSE*) is represented as the square root of the average of the squared differences between the predicted and actual values of a variable (Bishop & Nasrabadi, 2006), as shown in Eq. (9).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

4.3. Results

In this section, we discuss about the obtained results according to the experimental protocol discussed in Section 4.1.

5 https://www.nasdaq.com
6 http://colab.research.google.com/
7 https://scikit-learn.org/stable/
8 https://keras.io
9 https://www.tensorflow.org/

Table 5
Dataset characterization for stock forecasting task.

	Number	Training		Validation		Test	
		Sample	Date	Sample	Date	Sample	Date
Stock	417	480	01/11/2019–28/09/2021	140	29/09/2021–19/04/2022	70	20/04/2022–29/07/2022
Crypto	67	670	01/11/2019–28/09/2021	197	29/09/2021 – 19/04/2022	97	20/04/2022–31/07/2022

Table 6
Dataset characterization for advice task.

	Number	Training		Test	
		Sample	Date	Sample	Date
Stock	417	830	01/11/2019–29/07/2022	65	01/08/2022–31/10/2022
Crypto	67	964	01/11/2019–29/07/2022	92	01/08/2022–31/10/2022

Table 7
Comparison of expected Close price and Daily return on 70 days in the test set for each stock on the basis of average scores of MAE, MAPE, and RMSE metrics.

Stock	Close	DR	(Close-DR)
MAE	4.076	3.100	**0.976**
MAPE	0.028	0.021	**0.007**
RMSE	5.165	3.895	**1.270**

Table 8
Comparative analysis of forecasting module with and without seasonal data on the basis of average scores of MAE, MAPE, and RMSE metrics.

Stock	No Seasonal	Seasonal	Difference
MAE	3.100	2.874	**0.226**
MAPE	0.021	0.020	**0.001**
RMSE	3.895	3.574	**0.321**

Table 9
Comparative analysis of forecasting module with financial features and with technical indicator on the basis of average scores of MAE, MAPE, and RMSE metrics.

Stock	Financial	Indicators	Difference
MAE	2.873	2.484	**0.390**
MAPE	0.020	0.017	**0.003**
RMSE	3.574	3.257	**0.316**

Table 10
Comparative analysis of forecasting module with and without news data on the basis of average scores of MAE, MAPE, and RMSE metrics.

Stock	No News	News	Difference
MAE	2.484	2.469	**0.015**
MAPE	0.017	0.017	**4.943e−05**
RMSE	3.258	3.223	**0.035**

Table 11
Effectiveness performance of Forecasting module varying the window size on the basis of MAE, MAPE and RMSE metrics.

Window	MAE	MAPE	RMSE
2 Days	2.477	0.0174	3.238
3 Days	**2.464**	**0.0173**	**3.209**
4 Days	2.478	0.0173	3.247
5 Days	2.472	0.0173	3.227
6 Days	2.473	0.0174	3.228
7 Days	2.502	0.0174	3.249
8 Days	2.490	0.0174	3.257

4.3.1. Feature selection

The choice of the forecasting target plays a key role in designing the Forecasting Unit. Despite (Kumbure et al., 2022) usually predicting the *Closing price* of shares, we choose to predict the *Daily Return*, which appears to be a more accurate choice, as shown in Table 7. Specifically, we compare the predicted *Daily Returns* (DR) and *Closing price*, which is computed through Eq. (10) where $Open_i$ is the closing price, by using a Time Window of five days.

$$Close_i = Open_i + \frac{DR_i \cdot Open_i}{100} \quad \forall\ day\ i \tag{10}$$

Hence, we make a comparison between the expected Close price and Daily return by evaluating the prediction over 70 days in the test set for each stock. Table 7 shows the average score of MAE, MAPE and RMSE metrics, defined in Section 4.2.

It is worth noting in Table 7 that predicting *Daily Return* achieves the highest forecasting performances for all metrics w.r.t forecasting *Closing Price*. Hence, *Daily Return* has been chosen as the forecasting target for the Neural Networks in the proposed Framework.

Furthermore, we make other experiments to unveil *Daily Return* trends during fixed time intervals. Specifically, seasonal trends have been sought by grouping entries by: *Day* in *week* and *month* and *month* and *quarter* in *year* time period.

Observing the variability in the trend of *Daily Return* over the examined periods, it is reasonable to hypothesize that its value is dependent on the current Market Day in terms of weekday, day of the month, month and quarter. Hence, an experiment has been performed by adding these four categorical features to the dataset, whose results are shown in Table 8.

It is easy to note in Table 8 that the forecasting performances increase by adding these four features.

Another evaluation has been made by introducing the technical indicator, defined in Section 3.2 into the dataset to replace financial features. Table 9 shows the results, where it is possible to see that the use of Technical Indicators produces better forecasting performance with respect to Financial features for all metrics.

Finally, the last analysis aims to investigate how *News* data affects forecasting performances. In Table 10, it is worth noting that *News* data significantly improves effectiveness performance.

4.3.2. Hyperparameter tuning

This section describes the hyperparameter tuning to optimize the forecasting module (defined in Section 3.3) by varying two main parameters: (i) the number of historical stock samples to be fed as input to the forecast unit (*Time Window*) in the [2,8] days interval and (ii) *Neural Network* models selection by evaluating several deep learning models. Furthermore, we measure the forecasting error through the average score of MAE, MAPE, and RMSE metrics, defined in Section 4.2. Table 11 shows the forecasting performance varying Time Window size. The length of *Time Window* has a significant impact on forecasting performance because it indicates the number of days that the predictive model takes into account to compute *Daily Return* values.

It is worth noting that the 3-day time window provides better results than the other ones. This result is in agreement with the findings

Table 12
Architectural details about the baselines.

Layer (Type)	Output Shape	Param #
gru (GRU)	(None, 128)	74,880
dropout (Dropout)	(None, 128)	0
dense (Dense)	(None, 128)	16,512
dense_1 (Dense)	(None, 1)	129
Total params: 91,521 Trainable params: 91,521 Non-trainable params: 0		

(a) GRU Architecture

Layer (Type)	Output Shape	Param #
bidirectional (Bidirectional)	(None, 256)	198,656
dropout (Dropout)	(None, 256)	0
dense (Dense)	(None, 128)	32,896
dense_1 (Dense)	(None, 1)	129
Total params: 231,681 Trainable params: 231,681 Non-trainable params: 0		

(b) BI-LSTM Architecture

Layer (Type)	Output Shape	Param #
bidirectional (Bidirectional)	(None, 256)	149,760
dropout (Dropout)	(None, 256)	0
dense (Dense)	(None, 128)	32,896
dense_1 (Dense)	(None, 1)	129
Total params: 182,785 Trainable params: 182,785 Non-trainable params: 0		

(c) BI-GRU Architecture

in Weng et al. (2018), in which the authors show how using a time window approximately equal to the forecast horizon can improve the accuracy of stock market forecasts.

Furthermore, we compare different Deep Learning-based models (see Table 12) to identify the most suitable architecture for the forecasting task. Hence, we consider the Gated Recurrent Unit (GRU) (Cho et al., 2014), whose main cell (see Table 12(a)) is composed of two gates that enable a direct flow of information through the hidden state without using a separate memory cell as in the $LSTM$.

Another model is the Bidirectional $LSTM$ ($BI-LSTM$), which processes the input sequence bidirectionally through two separate hidden states (Schuster & Paliwal, 1997), that are, successively, combined to generate the final prediction. Table 12(b) shows the internal architecture of the examined model with the related number of parameters.

The last examined model is the Stacked $LSTM$ ($STKDLSTM$), which consists of several $LSTM$ layers (Graves, 2013), as shown in Table 12(c). Specifically, the output of each layer is fed as input into the next one. This allows the network to learn complex patterns in sequential data by constructing hierarchical representations of the input data. However, Stacked $LSTM$ networks are computationally expensive and require more training data to learn effectively. The stacking principle has also been applied to the other aforementioned models by incorporating two recurrent layers, resulting in the development of the following four neural networks: (i) Stacked LSTM ($STKDLSTM$), (ii) Stacked Bidirectional LSTM ($STKDBI-LSTM$), (iii) Stacked GRU ($STKDGRU$) and (iv) Stacked Bidirectional GRU ($STKDBI-GRU$).

Table 13 shows the comparison of the forecasting performance obtained for each Neural Network model. It is worth noting that $LSTM$ outperforms the other models. Hence, it has been chosen as the forecasting model for the proposed framework.

Table 13
Comparison of the proposed LSTM-based model w.r.t. seven baselines.

Model	MAE	MAPE	RMSE
LSTM	**2.458**	**0.0173**	**3.213**
BI-LSTM	2.476	0.0175	3.239
GRU	2.504	0.0178	3.254
BI-GRU	2.541	0.0177	3.315
STKD LSTM	2.459	0.0174	3.217
STKD BI-LSTM	2.507	0.0174	3.276
STKD GRU	2.515	0.0174	3.274
STKD BI-GRU	2.510	0.0176	3.275

Table 14
Effective analysis of the proposed LSTM-based model on some tech and booking stocks.

	MAE	MAPE	RMSE
AAPL	1.452	0.0169	2.681
MSFT	1.024	0.0155	2.311
GOOG	1.829	0.0172	3.012
BKNG	3.892	0.0193	3.956
ABNB	5.166	0.0214	5.105
Average	**2.6726**	**0.01806**	**3.413**

To provide a comprehensive view of the comparison, Fig. 4 summarize the obtained results of the comparison of the proposed model w.r.t. seven baselines in terms of MAE and RMSE. As shown in Banik et al. (2022), the most relevant measure is the RMSE because it penalizes high error values in forecasting analysis. This measure is relevant for our analysis because a lower RMSE value allows for a more realistic prediction of stock behavior, which, integrated into the proposed approach, provides more profitable stock advice.

In Table 14 we assess the effectiveness of the $LSTM$-based model on some tech and booking stocks. The proposed approach achieves the highest performance in identifying technology stock behavior while incurring a higher error rate on booking service provider one. This result is caused by the skewed trend of booking security due to the restrictions imposed during the COVID-19 pandemic.

4.3.3. Effectiveness analysis

In this section, we discuss the results obtained during the simulation process on the NASDAQ Stock and Cryptocurrency markets. In particular, daily advice has been generated for the 65 Open Market days from *01/08/2022* to *31/10/2022*. For each Open Market day, the proposed framework has been evaluated while the stocks are sorted in descending order by evaluation score to select the top 50 ones. This procedure enables to choose the securities showing the most predictable behavior on the current day. Hence, we make the forecast of the 50 selected stocks on the next day's *Daily Return*, also sorting them by expected *Daily Return* in descending order. Successively, the *Daily Advice* is generated by suggesting the 5 securities with the highest expected *Daily Return* for the next day. At the end of this process, we provide 65 daily investment advice, each one consists of five suggested stocks expected to exhibit an upward trend in the next day.

A daily trading strategy has been designed to evaluate the effectiveness of the advice. For each trading day, investment capital is equally divided among the 5 suggested securities by buying them at the *Opening Price* and selling them at the *Closing Price* without considering the related fees. Furthermore, we assumed that all available capital is daily invested to exploit compound interest to compute the economic gain obtained. In conclusion, the daily percentage change of the capital is provided by computing the *Daily Returns* main value of the traded securities.We achieve 64.62% as an accuracy score by obtaining positive and negative investment results, whose details are reported in Table 1 of the Appendix, for 42 and 23 days, respectively.

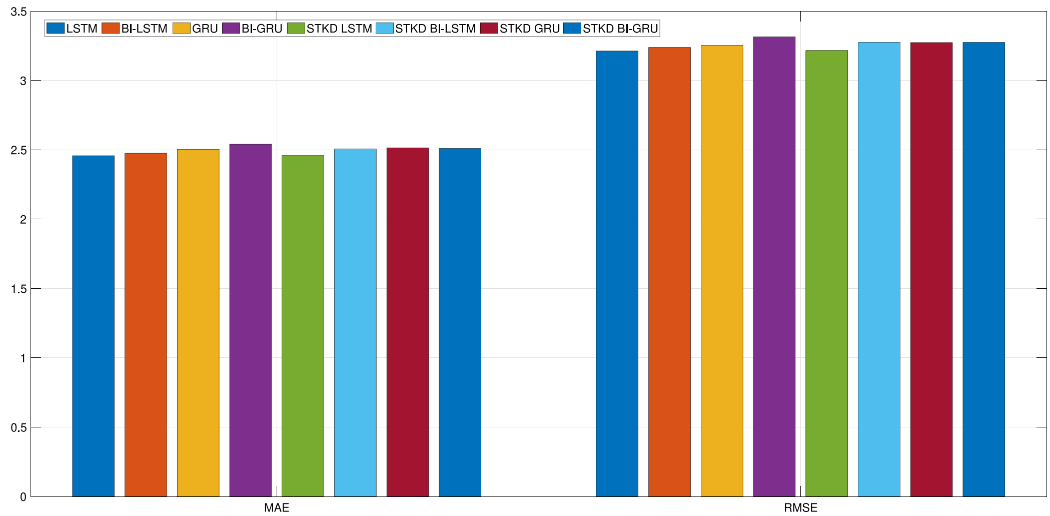

Fig. 4. Comparison of the proposed LSTM-based model w.r.t. seven baselines in terms of MAE and RMSE.

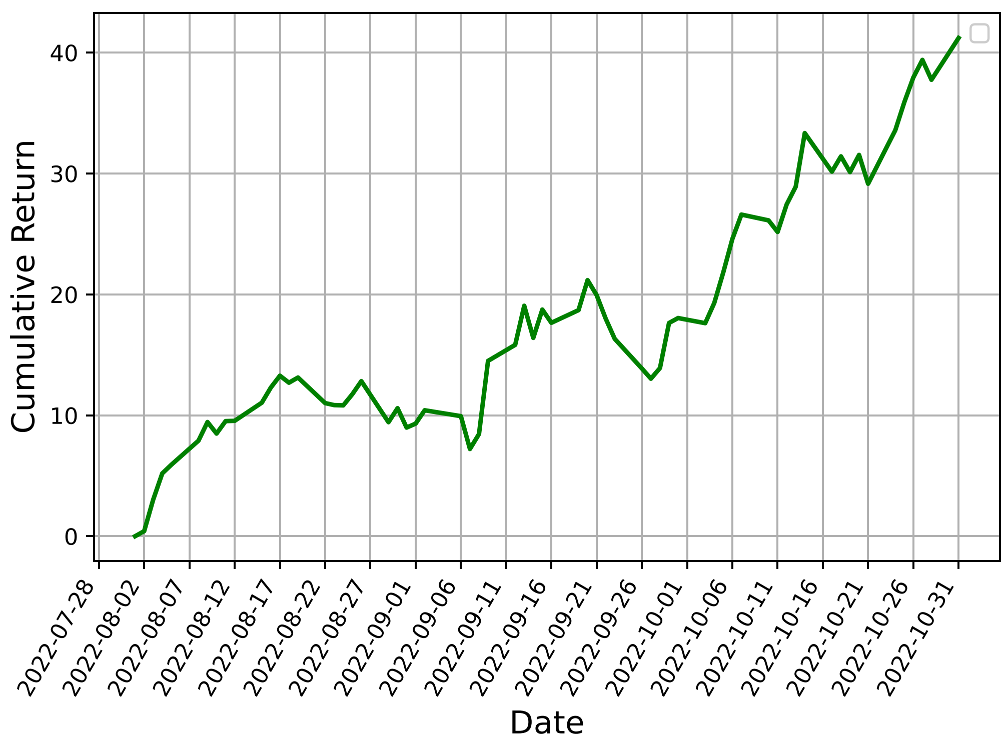

Fig. 5. Cumulative Percentage Economical Gain.

Let X_0 and DR_i be, respectively, the *Initial Capital* and the *Daily Return* at day i, the *Available Capital* at day k X_i can be calculated with Eq. (11):

$$X_i = X_{i-1} \cdot (1 + DR_i) \tag{11}$$

Hence, the *Cumulative Percentage Economical Gain* at day i G_i is obtained according to Eq. (12).

$$G_i = 100 \cdot \frac{X_i - X_0}{X_0} \tag{12}$$

Fig. 5 shows the value of the Cumulative Percentage Economic Gain over the three months simulation period. It is worth noting that the proposed framework produces advice leading to an economic gain of 41.21% of the initial capital.

Successively, the same simulation process has been further applied to a set of 67 cryptocurrencies. However, the simulation period spans over 92 trading days w.r.t. the 65 ones of the stock market, being the Cryptocurrency market open every day. We achieve 59.78% as an accuracy score by obtaining positive and negative investment results, whose details are reported in Table 2 of the Appendix, for 55 and 37 days, respectively.

Fig. 6 shows the value of the Cumulative Percentage Economic Gain over the three months simulation period. It is worth noting that the proposed framework produces advice leading to an economic gain of 39.38% of the initial capital.

4.3.4. Comparison w.r.t. state-of-the-art methods

In this section, we evaluate the performance of the proposed approach w.r.t. different state-of-the-art approaches on the test set. Firstly, we have predicted the stock behavior through a LSTM-based model. We have evaluated this task by using statistical metrics (i.e., MAE, MSE,

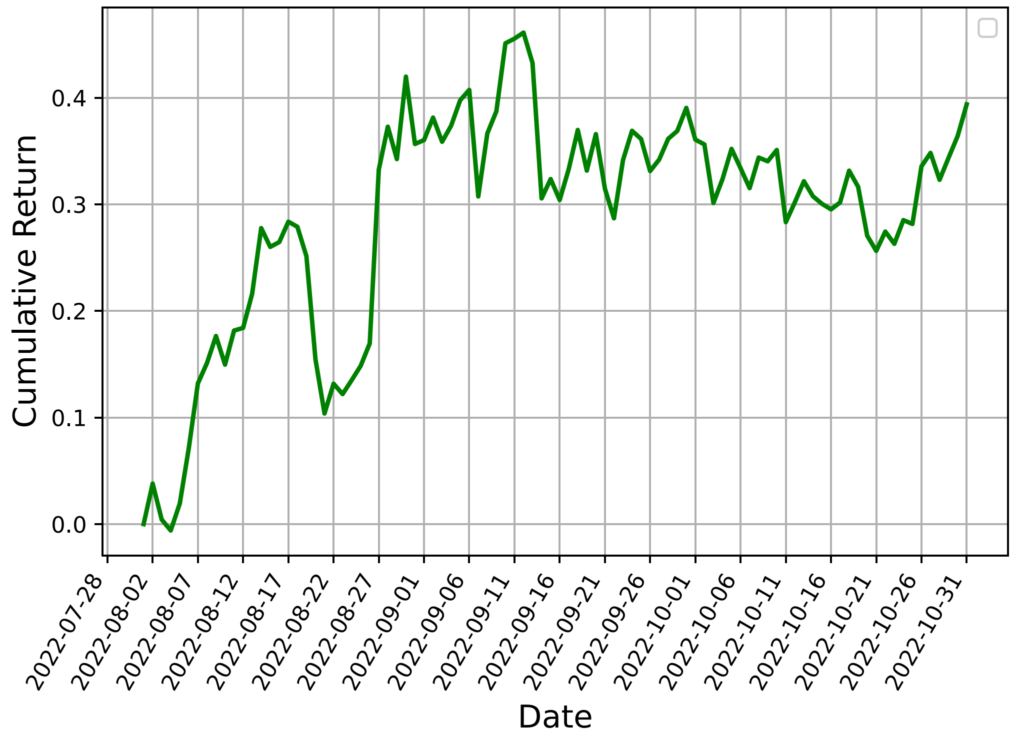

Fig. 6. Cryptocurrencies Cumulative Percentage Economical Gain.

Table 15
Effectiveness comparison of the proposed framework w.r.t. several state-of-the-art approaches.

	Stock Market			Cryptocurrencies		
Approaches	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Banik et al. (2022)	3.294	0.023	4.305	4.112	0.029	5.375
Hafiz et al. (2023)	3.228	0.023	4.219	4.030	0.028	5.268
Anbaee Farimani et al. (2022)	3.146	0.022	4.113	3.991	0.028	5.217
Chen et al. (2019)	2.950	0.021	3.856	3.870	0.027	5.059
Wang et al. (2022a)	3.490	0.025	4.562	4.172	0.029	5.454
Proposed	**2.728**	**0.019**	**3.566**	**3.749**	**0.026**	**4.900**

and RMSE) to assess the distance between the proposed model w.r.t. the real stock trend. Successively, we integrate the model's prediction into the stock selection strategy, which has been evaluated on the daily return metric, one of the most relevant return-based metrics. Table 15 shows the obtained results over two different markets (stock and cryptocurrencies). It is worth noticing that the proposed approach outperforms the baselines by considering different types of features enabling us to analyze the stock trend from different points of view. In fact, it is possible to see how the integration of contextual information (i.e., news and sentiment scores) with historical stock information allows for improved prediction of future stock performance. However, Table 15 shows a slight improvement of the proposed framework w.r.t. different baselines due to the dynamic and chaotic nature of the Cryptocurrency market (Jalan & Matkovskyy, 2023; Khosravi & Ghazani, 2023; Wang et al., 2022b). In fact, predicting cryptocurrency performance is challenging since it has no future market to take as a reference in contrast to the stock market and it is further difficult to reveal the potential factors that trigger cryptocurrency price movements, as shown in Dolatsara et al. (2022).

Finally, we compare the proposed advice strategy with respect to the one designed by Ghosh et al. (2022) over the entire test shown in Table 6. Table 16 shows the outcome of the comparison on both the market (i.e., stock and cryptocurrencies) in terms of cumulative return.

Table 16
Effectiveness comparison of the proposed advice strategy w.r.t. the buy & hold and the one designed by Ghosh et al. (2022).

	Cumulative Return	
	Stock	Crypto
Buy & hold strategy	28.31%	26.42%
Ghosh et al. (2022)	35.46%	33.81%
Proposal	**41.21%**	**39.38%**

It is easy to note that the proposed approach outperforms the strategy proposed by (buy & hold) of 16.22% (45.05%) and 16.47% (49.05%) for stock market and cryptocurrencies, respectively. Although the strategy designed by Ghosh et al. (2022) aims to combine different features, it only relies on historical financial prices without considering external factors and possible financial fundamental analysis.

5. Conclusion

In the last years, the pervasive diffusion of Artificial Intelligence in the finance domain transformed different services, including stock forecasting. In particular, the stock market has become an investment channel for large profits, prompting researchers and practitioners to design and apply Artificial Intelligence-based models in order to improve effectiveness in predicting stock performance.

Despite different Artificial Intelligence-based approaches have been proposed for stock forecasting, they are only focused on news content or sentiment without fundamental features (Anbaee Farimani et al., 2022; Chen et al., 2019) and vice versa (Banik et al., 2022). In turn, other approaches rely on the handmade rules (Ozcalici & Bumin, 2022) or ones based on technical indicators (Banik et al., 2022) for providing advice without considering contextual information that can strongly affect the stock market.

In this paper, we developed a $LSTM$ framework for daily stock market advice to provide investment suggestions that could lead to economic gain. In particular, the developed framework relies on $LSTM$-based models capable of making predictions from heterogeneous data incorporating financial information, investor sentiment, and market seasonality. Furthermore, the proposed *Heuristic Stocks Selection* algorithm allows the framework to choose stocks showing the most predictable behavior on the previous open market day.

To evaluate its effectiveness, the proposed framework has first been evaluated in a three-month simulation on over 400 NASDAQ stocks, achieving an economic gain of 41.21% of the initial capital, despite the downward trend of the NASDAQ Stock Market in the quarter under review. In addition, it has also been tested on the Cryptocurrency Market showing an economic gain of 39.38%. We further compared the proposed framework with several state-of-the-art approaches, showing how it outperforms the baseline in both markets. Furthermore, the low memory space occupation and training time required from the AI-based forecasting unit is mainly due to the amount of input information, that is related to 3 days time window (as shown in Section 4.3.2 during the hyperparameter tuning process).

However, the proposed approach suffers from some limitations, that will drive our future research activities on this topic. The proposed strategy does not take into account the transaction costs although they may integrate into the stock selection algorithm. Furthermore, the advice strategy does not consider possible risks of the investment, which may be taken into account through the Sharpe Ratio and/or D-ratio to investigate the distributed returns.

Future works will be devoted to increasing the dataset size in terms of the number of markets and years. They will be further focused on designing more complex and efficient trading strategies that can increase the financial gain achieved by following the generated investment advice. In addition, we will investigate how transaction costs can be integrated into the proposed advice strategy as weights to steer the selection process. Finally, innovative approaches that incorporate the real-world relationships between financial products will certainly be explored in order to exploit the correlations between them.

Funding

This work is supported by the Italian Ministry of Education, University and Research (MIUR) within the PRIN2022—ISALDI: Interpretable Stock Analysis Leveraging Deep multImodal models (CUP: E53D23008150006).

CRediT authorship contribution statement

Fausto Ricchiuti: Conception and design of study, Acquisition of data, Analysis and interpretation of data, Writing – original draft. **Giancarlo Sperlí:** Conception and design of study, Acquisition of data, Analysis and interpretation of data, Writing – original draft.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Appendix

The proposed framework has been evaluated on two different scenarios: stock and cryptocurrency markets. In this appendix, we detail the obtained results in terms of aggregated accuracy over the entire test set and Daily Return. We choose DR as a relevant metric to assess the proposed advice strategy because it is one of the most used metrics to evaluate the profitability of a model or strategy (Behera et al., 2023; Olorunnimbe & Viktor, 2022).

In the stock market, we obtain that the investment results are positive for 42 days of Trading and negative for 23, achieving an *Accuracy* of 64.62%. Furthermore, Table 17 shows the obtained results in terms of daily return for each day in the test set.

In turn, we achieve an accuracy of 59.78% on the Cryptocurrency market, being the investment results positive and negative for 55 and 37 days, respectively. Furthermore, Table 18 shows the results obtained by considering 67 Cryptocurrencies in terms of daily return for each day in the test set.

Table 17
Daily outcome of the proposed advice on Stock market evaluated in terms of daily return (DR) for each day in the test set.

Daily Advice		
Date	Traded Stocks	DR
01-08-2022	NFE - SWKS - OMCL - MEDP - EXPE	0.42%
02-08-2022	APPN - RMBS - TNDM - SAIA - AMZN	2.6%
03-08-2022	SPWR - PCRX - AEIS - ALNY - SIRI	2.1%
04-08-2022	ETSY - ITCI - WTFC - GH - INDB	0.68%
05-08-2022	DNLI - EEFT - ALTR - LSXMK - GOOG	1.88%
08-08-2022	PEGA - NVDA - JKHY - PDCE - WAFD	1.43%
09-08-2022	UHAL - CALM - SSNC - RUSHA - POOL	−0.87%
10-08-2022	PCTY - PYPL - TRMB - TENB - APA	0.95%
11-08-2022	NXST - WEN - LBRDA - SPSC - STLD	0.02%
12-08-2022	SYNA - WBA - NXST - VRTX - CHX	1.37%
15-08-2022	FIZZ - MASI - PCH - BIIB - BHF	1.13%
16-08-2022	SLM - PEGA - BCPC - FISV - NTRS	0.87%
17-08-2022	WWD - FOX - PENN - NSIT - OKTA	−0.51%
18-08-2022	OLLI - CYTK - PCTY - CRVL - TER	0.38%
19-08-2022	DXCM - MTCH - SFNC - MAR - WING	−1.87%
22-08-2022	OMCL - FIBK - MRTX - MXL - CHRD	−0.15%
23-08-2022	ARVN - QDEL - WERN - PCRX - VRRM	−0.02%
24-08-2022	FOLD - PCTY - TSCO - ZBRA - ANSS	0.83%
25-08-2022	CHDN - LAMR - KRTX - LKQ - HBAN	0.97%
26-08-2022	MIDD - SBUX - UFPI - EYE - SGEN	−3.01%
29-08-2022	PTEN - AZPN - SFM - EA - FANG	1.06%
30-08-2022	COKE - MEDP - MORN - SAIA - EEFT	−1.45%
31-08-2022	WDAY - ON - CWST - MRTX - FANG	0.3%
01-09-2022	CHTR - LAMR - VNOM - LOGI - ROST	1.01%
02-09-2022	OLED - CHDN - GH - WEN - SYNA	−0.44%
06-09-2022	PCRX - MGEE - CYTK - FRME - IEP	−2.47%
07-09-2022	NDSN - MXL - ISRG - RMBS - TW	1.15%
08-09-2022	CAR - DOCU - TENB - ISEE - TMUS	5.58%
09-09-2022	CSGP - MXL - ACIW - LAMR - PANW	1.16%
12-09-2022	VIAV - FULT - NTCT - EEFT - CRUS	2.79%
13-09-2022	FIVN - DNLI - EXLS - OLED - MRNA	−2.23%
14-09-2022	PRTA - AAL - MTSI - FOXA - KLAC	2.01%
15-09-2022	ISEE - ABMD - IBOC - LSCC - ZS	−0.92%
16-09-2022	ON - GLPI - BECN - OKTA - MASI	0.89%
19-09-2022	ON - ACIW - ENTG - SFM - CWST	2.09%
20-09-2022	APPF - ISRG - TMUS - META - AMZN	−1.02%
21-09-2022	NBIX - CATY - PYPL - SLM - HBAN	−1.62%
22-09-2022	SYNA - ACIW - FOXF - RMBS - IPGP	−1.42%
23-09-2022	ARLP - PPC - IBOC - EXEL - FOLD	−2.11%
26-09-2022	WEN - ORLY - FISV - ZBRA - OMCL	−0.74%
27-09-2022	SWAV - EWBC - ITCI - CVLT - TXRH	0.78%
28-09-2022	MTCH - MAR - ETSY - LFUS - BMRN	3.27%
29-09-2022	AXNX - TENB - ZBRA - AXON - SLM	0.35%
30-09-2022	ITCI - FIVN - ZM - LITE - FOXA	−0.36%
03-10-2022	UAL - SBAC - LSXMK - BECN - IRTC	1.42%
04-10-2022	STAA - GH - CG - CYTK - RUN	2.13%
05-10-2022	PLXS - RMBS - FTNT - CHX - MDLZ	2.26%
06-10-2022	LPLA - FULT - CYTK - TNDM - NDAQ	1.62%
07-10-2022	ITCI - PRTA - FWRD - ACIW - IBTX	−0.38%
10-10-2022	MXL - ZM - EXC - CELH - NTAP	−0.76%
11-10-2022	STLD - SANM - CASY - SMPL - BLKB	1.82%
12-10-2022	EYE - ORLY - GH - CACC - WAFD	1.15%
13-10-2022	PAGP - GNTX - UHAL - SGEN - SFNC	3.44%

(*continued on next page*)

Table 17 (*continued*).

Daily Advice		
Date	Traded Stocks	DR
14-10-2022	RGLD - ARLP - NFE - PINC - AXNX	−2.39%
17-10-2022	PTCT - GNTX - PINC - LAMR - FOXA	0.97%
18-10-2022	BCRX - AZPN - SAIA - TCBI - UFPI	−0.99%
19-10-2022	CSGP - NFE - LRCX - TFSL - CSX	1.09%
20-10-2022	GT - CHDN - FFIN - PACW - IBOC	−1.81%
21-10-2022	TER - LSCC - AAPL - MXL - HALO	3.42%
24-10-2022	ENPH - EYE - REGN - SAIA - ABCB	1.74%
25-10-2022	LRCX - PARA - LSXMA - SAIA - WWD	1.51%
26-10-2022	KLAC - WERN - PPBI - VNOM - LSTR	1.05%
27-10-2022	VRNS - AXON - NTNX - AMKR - CALM	−1.18%
28-10-2022	RGEN - TTEK - RMBS - ERIE - CTAS	2.51%
31-10-2022	TENB - NTNX - CSGP - ODFL - LANC	1.32%

Table 18
Daily outcome of the proposed advice strategy on Cryptocurrency market evaluated in terms of daily return (DR) for each day in the test set.

Daily Cryptocurrencies Advice		
Date	Traded Cryptos	DR
01-08-2022	NEBL - BNT - DASH - CVC - LTC	3.8%
02-08-2022	ATOM - WAVES - EOS - BTT - CVC	−3.25%
03-08-2022	TRX - ETC - BTG - XTZ - FTT	−1.02%
04-08-2022	FTM - DASH - BNT - EOS - ATOM	2.55%
05-08-2022	SNX - DASH - XMR - ALGO - FTM	5.12%
06-08-2022	STORJ - DOGE - NEBL - LTC - HBAR	5.64%
07-08-2022	XTZ - ATOM - SC - CVC - MANA	1.68%
08-08-2022	ANT - SC - BCH - XLM - FTM	2.21%
09-08-2022	CHZ - TRX - ATOM - QNT - ANT	−2.28%
10-08-2022	ZEC - ALGO - FTT - VGX - ANT	2.78%
11-08-2022	STORJ - XEM - NEO - MATIC - ALGO	0.21%
12-08-2022	BCH - META - NEBL - CEL - BSV	2.71%
13-08-2022	GNO - HBAR - CEL - SC - ZEN	5.06%
14-08-2022	LRC - ATOM - CEL - CRO - TRX	−1.38%
15-08-2022	ZEC - ALGO - DASH - CHZ - TRX	0.36%
16-08-2022	ZEC - ATOM - NEO - ANT - CHZ	1.51%
17-08-2022	CHZ - FTT - DASH - SC - HBAR	−0.37%
18-08-2022	XEM - REP - LINK - CHZ - CEL	−2.16%
19-08-2022	STORJ - MATIC - NEBL - REP - LINK	−7.73%
20-08-2022	ATOM - ANT - ZEN - FTT - CEL	−4.4%
21-08-2022	FIL - OMG - FTM - META - ZEC	2.53%
22-08-2022	ANT - WAVES - XLM - DASH - XEM	−0.86%
23-08-2022	STORJ - KCS - BSV - MATIC - ONE	1.16%
24-08-2022	MANA - REP - NEBL - CHZ - SNX	1.19%
25-08-2022	XTZ - HBAR - ZEC - DASH - OMG	1.83%
26-08-2022	ANT - NEBL - DASH - GNO - VGX	13.94%
27-08-2022	OMG - ALGO - SNX - ADA - LINK	3.02%
28-08-2022	OMG - FTT - ALGO - REP - XTZ	−2.21%
29-08-2022	FTM - MTL - BNT - REP - MKR	5.76%
30-08-2022	NEBL - SNX - XEM - ALGO - ANT	−4.45%
31-08-2022	NEBL - BNB - MIOTA - STORJ - ADA	0.27%
01-09-2022	NEBL - STORJ - LINK - BAT - DOGE	1.55%
02-09-2022	BSV - HBAR - XMR - NEBL - MTL	−1.64%
03-09-2022	BSV - CEL - LINK - XTZ - FTM	1.11%
04-09-2022	ATOM - ZEC - XMR - MIOTA - ONE	1.74%
05-09-2022	BCH - CVC - REP - WAVES - ATOM	0.69%
06-09-2022	DASH - BAT - ARK - KCS - LRC	−7.1%
07-09-2022	FIL - REP - ATOM - HBAR - FTT	4.52%
08-09-2022	ALGO - BTT - STORJ - WAVES - CRO	1.52%
09-09-2022	LTC - SNX - MATIC - BNB - OMG	4.6%
10-09-2022	CVC - BNB - ZEN - ZEC - ADA	0.3%
11-09-2022	QNT - ALGO - FTM - ETH - BNB	0.39%
12-09-2022	VGX - NEO - LINK - CHZ - CRO	−1.96%

(*continued on next page*)

Table 18 (*continued*).

Daily Cryptocurrencies Advice		
Date	Traded Cryptos	DR
13-09-2022	MATIC - FIL - HBAR - XEM - FTM	−8.86%
14-09-2022	ANT - STORJ - BSV - QNT - CRO	1.39%
15-09-2022	BNB - STORJ - VGX - NEO - BNT	−1.48%
16-09-2022	MATIC - NEO - ATOM - LRC - XEM	2.22%
17-09-2022	ZEN - CVC - FTT - MIOTA - LRC	2.75%
18-09-2022	ATOM - ZEN - BCH - XTZ - ONE	−2.78%
19-09-2022	SNX - GNO - FTT - LRC - DASH	2.57%
20-09-2022	SNX - MATIC - FIL - ADA - REP	−3.74%
21-09-2022	FTT - NEO - OMG - MTL - QNT	−2.12%
22-09-2022	FIL - CVC - TRX - STORJ - DOGE	4.24%
23-09-2022	CVC - MATIC - MTL - QNT - TRX	2.05%
24-09-2022	XRP - QNT - ARK - XVG - DASH	−0.56%
25-09-2022	LRC - HBAR - BTT - SNX - STORJ	−2.21%
26-09-2022	BTT - NEO - BCH - LINK - MTL	0.81%
27-09-2022	META - CEL - BNT - NEO - DASH	1.45%
28-09-2022	STORJ - BTT - XTZ - OMG - MATIC	0.54%
29-09-2022	OMG - FTM - GNO - XLM - LTC	1.58%
30-09-2022	BTT - CEL - XEM - LTC - CRO	−2.14%
01-10-2022	FIL - XLM - DOGE - XVG - ATOM	−0.33%
02-10-2022	XLM - CHZ - MTL - OMG - META	−4.05%
03-10-2022	ALGO - MANA - CEL - META - XEM	1.72%
04-10-2022	XRP - BNT - META - HBAR - FTM	2.14%
05-10-2022	QNT - SNX - FTT - KCS - XTZ	−1.34%
06-10-2022	HBAR - LTC - ANT - CVC - ONE	−1.4%
07-10-2022	CHZ - FIL - NEBL - KCS - XRP	2.18%
08-10-2022	XRP - MATIC - XEM - MTL - BAT	−0.26%
09-10-2022	VGX - BNT - MTL - XVG - OMG	0.79%
10-10-2022	TRX - CHZ - OMG - LRC - NEBL	−5.01%
11-10-2022	XMR - WAVES - ONE - CVC - FUN	1.45%
12-10-2022	HBAR - SNX - ALGO - TRX - FIL	1.52%
13-10-2022	HBAR - META - XEM - LINK - FUN	−1.07%
14-10-2022	FTM - META - ETC - ARK - GNO	−0.54%
15-10-2022	ADA - XEM - ATOM - ANT - XLM	−0.4%
16-10-2022	FIL - HBAR - XEM - CHZ - KCS	0.5%
17-10-2022	MATIC - BNB - CVC - ADA - KCS	2.29%
18-10-2022	ATOM - BAT - FTT - BSV - BNB	−1.14%
19-10-2022	LRC - KCS - MANA - ALGO - MTL	−3.48%
20-10-2022	FIL - STORJ - XTZ - XMR - XEM	−1.12%
21-10-2022	MANA - BCH - MTL - XTZ - CVC	1.44%
22-10-2022	ZEN - ADA - BNT - QNT - HBAR	−0.9%
23-10-2022	LRC - XVG - LINK - CVC - HBAR	1.76%
24-10-2022	GNO - ARK - HBAR - BNB - VGX	−0.28%
25-10-2022	MANA - ALGO - VGX - NEBL - BTG	4.22%
26-10-2022	NEBL - HBAR - MATIC - XRP - QNT	0.94%
27-10-2022	NEBL - ETH - ANT - BNB - QNT	−1.87%
28-10-2022	CEL - STORJ - FTT - BSV - HBAR	1.57%
29-10-2022	LRC - MANA - VGX - ADA - XRP	1.51%
30-10-2022	NEBL - FUN - BNB - WAVES - FTT	2.18%
31-10-2022	BNB - XEM - VGX - LINK - CVC	0.82%

References

de A. Araújo, R., & Ferreira, T. A. (2013). A morphological-rank-linear evolutionary method for stock market prediction. *Information Sciences*, *237*, 3–17. http://dx.doi.org/10.1016/j.ins.2009.07.007.

Abdi, H. (2007). The Kendall rank correlation coefficient. In *Encyclopedia of measurement and statistics* (pp. 508–510). Thousand Oaks, CA: Sage.

Agarwal, S., Kumar, S., & Goel, U. (2019). Stock market response to information diffusion through internet sources: A literature review. *International Journal of Information Management*, *45*, 118–131. http://dx.doi.org/10.1016/j.ijinfomgt.2018.11.002.

Ahelegbey, D. F., Cerchiello, P., & Scaramozzino, R. (2022). Network based evidence of the financial impact of Covid-19 pandemic. *International Review of Financial Analysis*, *81*, Article 102101. http://dx.doi.org/10.1016/j.irfa.2022.102101.

Ahmed, S., Chakrabortty, R. K., Essam, D. L., & Ding, W. (2022). Poly-linear regression with augmented long short term memory neural network: Predicting time series data. *Information Sciences*, *606*, 573–600. http://dx.doi.org/10.1016/j.ins.2022.05.078.

Anbaee Farimani, S., Vafaei Jahan, M., Milani Fard, A., & Tabbakh, S. R. K. (2022). Investigating the informativeness of technical indicators and news sentiment in financial market price prediction. *Knowledge-Based Systems*, *247*, Article 108742. http://dx.doi.org/10.1016/j.knosys.2022.108742.

Ashtiani, M. N., & Raahemi, B. (2023). News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. *Expert Systems with Applications*, *217*, Article 119509. http://dx.doi.org/10.1016/j.eswa.2023.119509.

Azarnejad, A., & Khaloozadeh, H. (2022). Stock return system identification and multiple adaptive forecast algorithm for price trend forecasting. *Expert Systems with Applications*, *198*, Article 116685. http://dx.doi.org/10.1016/j.eswa.2022.116685.

Bai, J., Guo, J., Sun, B., Guo, Y., Bao, Q., & Xiao, X. (2023). Intelligent forecasting model of stock price using neighborhood rough set and multivariate empirical mode decomposition. *Engineering Applications of Artificial Intelligence*, *122*, Article 106106. http://dx.doi.org/10.1016/j.engappai.2023.106106.

Banik, S., Sharma, N., Mangla, M., Mohanty, S. N., & S., S. (2022). LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems*, *239*, Article 107994. http://dx.doi.org/10.1016/j.knosys.2021.107994.

Behera, J., Pasayat, A. K., Behera, H., & Kumar, P. (2023). Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. *Engineering Applications of Artificial Intelligence*, *120*, Article 105843. http://dx.doi.org/10.1016/j.engappai.2023.105843.

Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, Y., Maddix, D., Turkmen, C., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L., Aubet, F. X., Callot, L., & Januschowski, T. (2022). Deep learning for time series forecasting: Tutorial and literature survey. *ACM Computing Surveys*, *55*(6), http://dx.doi.org/10.1145/3533382.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning: vol. 4*, (4), Springer.

Cagliero, L., Garza, P., Attanasio, G., & Baralis, E. (2020). Training ensembles of faceted classification models for quantitative stock trading. *Computing*, *102*, 1213–1225. http://dx.doi.org/10.1007/s00607-019-00776-7.

Cao, L. (2022). AI in finance: Challenges, techniques, and opportunities. *ACM Computing Surveys*, *55*(3), http://dx.doi.org/10.1145/3502289.

Cervelló-Royo, R., & Guijarro, F. (2020). Forecasting stock market trend: A comparison of machine learning algorithms. *Finance, Markets and Valuation*, *6*(1), 37–49. http://dx.doi.org/10.46503/NLUF8557.

Chande, T. S. (2001). *Beyond technical analysis: How to develop and implement a winning trading system: vol. 101*, John Wiley & Sons.

Chen, W., Chen, B., & Cai, X. (2023). Forecasting China's stock market risk under the background of the stock connect programs. *Soft Computing*, 1–17. http://dx.doi.org/10.1007/s00500-023-08496-z.

Chen, W., Jiang, M., Zhang, W. G., & Chen, Z. (2021). A novel graph convolutional feature based convolutional neural network for stock trend prediction. *Information Sciences*, *556*, 67–94. http://dx.doi.org/10.1016/j.ins.2020.12.068.

Chen, M. Y., Liao, C. H., & Hsieh, R. P. (2019). Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach. *Computers in Human Behavior*, *101*, 402–408. http://dx.doi.org/10.1016/j.chb.2019.03.021.

Chen, X., Ma, X., Wang, H., Li, X., & Zhang, C. (2022). A hierarchical attention network for stock prediction based on attentive multi-view news learning. *Neurocomputing*, *504*, 1–15. http://dx.doi.org/10.1016/j.neucom.2022.06.106.

Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean–variance portfolio optimization using machine learning-based stock price prediction. *Applied Soft Computing*, *100*, Article 106943. http://dx.doi.org/10.1016/j.asoc.2020.106943.

Cheng, C. H., & Yang, J. H. (2018). Fuzzy time-series model based on rough set rule induction for forecasting stock price. *Neurocomputing*, *302*, 33–45. http://dx.doi.org/10.1016/j.neucom.2018.04.014.

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.

Choi, M., Lee, H. J., Park, S. H., Jeon, S. W., & Cho, S. (2024). Stock price momentum modeling using social media data. *Expert Systems with Applications*, *237*, Article 121589. http://dx.doi.org/10.1016/j.eswa.2023.121589.

Dessain, J. (2022). Machine learning models predicting returns: Why most popular performance metrics are misleading and proposal for an efficient metric. *Expert Systems with Applications*, *199*, Article 116970. http://dx.doi.org/10.1016/j.eswa.2022.116970.

Dezhkam, A., & Manzuri, M. T. (2023). Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert–Huang transform. *Engineering Applications of Artificial Intelligence*, *118*, Article 105626. http://dx.doi.org/10.1016/j.engappai.2022.105626.

Ding, S., & Li, R. (2021). Forecasting the sales and stock of electric vehicles using a novel self-adaptive optimized grey model. *Engineering Applications of Artificial Intelligence*, *100*, Article 104148. http://dx.doi.org/10.1016/j.engappai.2020.104148.

Dolatsara, H. A., Kibis, E., Caglar, M., Simsek, S., Dag, A., Dolatsara, G. A., & Delen, D. (2022). An interpretable decision-support systems for daily cryptocurrency trading. *Expert Systems with Applications*, *203*, Article 117409. http://dx.doi.org/10.1016/j.eswa.2022.117409.

Dong, H., Ren, J., Padmanabhan, B., & Nickerson, J. V. (2022). How are social and mass media different in relation to the stock market? A study on topic coverage and predictive value. *Information & Management*, *59*(2), Article 103588. http://dx.doi.org/10.1016/j.im.2021.103588.

Efendi, R., Arbaiy, N., & Deris, M. M. (2018). A new procedure in stock market forecasting based on fuzzy random auto-regression time series model. *Information Sciences*, *441*, 113–132. http://dx.doi.org/10.1016/j.ins.2018.02.016.

Gangopadhyay, S., & Majumder, P. (2023). Text representation for direction prediction of share market. *Expert Systems with Applications*, *211*, Article 118472. http://dx.doi.org/10.1016/j.eswa.2022.118472.

Ge, W., Lalbakhsh, P., Isai, L., Lenskiy, A., & Suominen, H. (2022). Neural network–based financial volatility forecasting: A systematic review. *ACM Computing Surveys*, *55*(1), http://dx.doi.org/10.1145/3483596.

Ghosh, P., Neufeld, A., & Sahoo, J. K. (2022). Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. *Finance Research Letters*, *46*, Article 102280. http://dx.doi.org/10.1016/j.frl.2021.102280.

Gjerstad, P., Meyn, P. F., Molnár, P., & Næss, T. D. (2021). Do president trump's tweets affect financial markets? *Decision Support Systems*, *147*, Article 113577. http://dx.doi.org/10.1016/j.dss.2021.113577.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256). JMLR Workshop and Conference Proceedings.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Graves, A. (2013). Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.

Hafiz, F., Broekaert, J., La Torre, D., & Swain, A. (2023). Co-evolution of neural architectures and features for stock market forecasting: A multi-objective decision perspective. *Decision Support Systems*, Article 114015. http://dx.doi.org/10.1016/j.dss.2023.114015.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). Survey on categorical data for neural networks. *Journal of Big Data*, *7*(1), 1–41. http://dx.doi.org/10.1186/s40537-020-00305-w.

Haq, A. U., Zeb, A., Lei, Z., & Zhang, D. (2021). Forecasting daily stock trend using multi-filter feature selection and deep learning. *Expert Systems with Applications*, *168*, Article 114444. http://dx.doi.org/10.1016/j.eswa.2020.114444.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, *124*, 226–251. http://dx.doi.org/10.1016/j.eswa.2019.01.012.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2023). Practical machine learning: Forecasting daily financial markets directions. *Expert Systems with Applications*, *233*, Article 120840. http://dx.doi.org/10.1016/j.eswa.2023.120840.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735.

Htun, H. H., Biehl, M., & Petkov, N. (2023). Survey of feature selection and extraction techniques for stock market prediction. *Financial Innovation*, *9*(1), 26. http://dx.doi.org/10.1186/s40854-022-00441-7.

Huang, W. C., Chen, C. T., Lee, C., Kuo, F. H., & Huang, S. H. (2023). Attentive gated graph sequence neural network-based time-series information fusion for financial trading. *Information Fusion*, *91*, 261–276. http://dx.doi.org/10.1016/j.inffus.2022.10.006.

Jalan, A., & Matkovskyy, R. (2023). Systemic risks in the cryptocurrency market: Evidence from the FTX collapse. *Finance Research Letters*, *53*, Article 103670. http://dx.doi.org/10.1016/j.frl.2023.103670.

Ji, G., Yu, J., Hu, K., Xie, J., & Ji, X. (2022). An adaptive feature selection schema using improved technical indicators for predicting stock price movements. *Expert Systems with Applications*, *200*, Article 116941. http://dx.doi.org/10.1016/j.eswa.2022.116941.

Jin, K., Sun, S., Li, H., & Zhang, F. (2022). A novel multi-modal analysis model with baidu search index for subway passenger flow forecasting. *Engineering Applications of Artificial Intelligence*, *107*, Article 104518. http://dx.doi.org/10.1016/j.engappai.2021.104518.

Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, *178*, Article 115019. http://dx.doi.org/10.1016/j.eswa.2021.115019.

Kamara, A. F., Chen, E., & Pan, Z. (2022). An ensemble of a boosted hybrid of deep learning models and technical analysis for forecasting stock prices. *Information Sciences*, *594*, 1–19. http://dx.doi.org/10.1016/j.ins.2022.02.015.

Kehinde, T., Chan, F. T., & Chung, S. (2023). Scientometric review and analysis of recent approaches to stock market forecasting: Two decades survey. *Expert Systems with Applications*, *213*, Article 119299. http://dx.doi.org/10.1016/j.eswa.2022.119299.

Khosravi, M., & Ghazani, M. M. (2023). Novel insights into the modeling financial time-series through machine learning methods: Evidence from the cryptocurrency market. *Expert Systems with Applications*, *234*, Article 121012. http://dx.doi.org/10.1016/j.eswa.2023.121012.

Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications*, *103*, 25–37. http://dx.doi.org/10.1016/j.eswa.2018.03.002.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, *197*, Article 116659. http://dx.doi.org/10.1016/j.eswa.2022.116659.

Li, Q., Chen, Y., Wang, J., Chen, Y., & Chen, H. (2018). Web media and stock markets : A survey and future directions from a big data perspective. *IEEE Transactions on Knowledge and Data Engineering*, *30*(2), 381–399. http://dx.doi.org/10.1109/TKDE.2017.2763144.

Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management*, *57*(5), Article 102212. http://dx.doi.org/10.1016/j.ipm.2020.102212.

Ma, X., Zhao, T., Guo, Q., Li, X., & Zhang, C. (2022). Fuzzy hypergraph network for recommending top-K profitable stocks. *Information Sciences*, *613*, 239–255. http://dx.doi.org/10.1016/j.ins.2022.09.010.

Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., Selim, M. M., & Muhammad, K. (2020). A local and global event sentiment based efficient stock exchange forecasting using deep learning. *International Journal of Information Management*, *50*, 432–451. http://dx.doi.org/10.1016/j.ijinfomgt.2019.07.011.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning* (pp. 807–814).

Nazareth, N., & Ramana Reddy, Y. V. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications*, *219*, Article 119640. http://dx.doi.org/10.1016/j.eswa.2023.119640.

Olorunnimbe, K., & Viktor, H. (2022). Deep learning in the stock market—A systematic survey of practice, backtesting, and applications. *Artificial Intelligence Review*, 1–53. http://dx.doi.org/10.1007/s10462-022-10226-0.

Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. *Applied Soft Computing*, *93*, Article 106384. http://dx.doi.org/10.1016/j.asoc.2020.106384.

Ozcalici, M., & Bumin, M. (2022). Optimizing filter rule parameters with genetic algorithm and stock selection with artificial neural networks for an improved trading: The case of borsa Istanbul. *Expert Systems with Applications*, *208*, Article 118120. http://dx.doi.org/10.1016/j.eswa.2022.118120.

Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, *33*(6), 497–505. http://dx.doi.org/10.1016/j.omega.2004.07.024.

Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems with Applications*, *115*, 635–655. http://dx.doi.org/10.1016/j.eswa.2018.08.003.

Park, S., & Yang, J. S. (2024). Machine learning modeling to forecast uncertainty between capital sudden stop and boom. *Expert Systems with Applications*, *237*, Article 121662. http://dx.doi.org/10.1016/j.eswa.2023.121662.

Picasso, A., Merello, S., Ma, Y., Oneto, L., & Cambria, E. (2019). Technical analysis and sentiment embeddings for market trend prediction. *Expert Systems with Applications*, *135*, 60–70. http://dx.doi.org/10.1016/j.eswa.2019.06.014.

Ronaghi, F., Salimibeni, M., Naderkhani, F., & Mohammadi, A. (2022). COVID19-HPSMP: COVID-19 adopted hybrid and parallel deep information fusion framework for stock price movement prediction. *Expert Systems with Applications*, *187*, Article 115879. http://dx.doi.org/10.1016/j.eswa.2021.115879.

Sampaio, J. M., & Morettin, P. A. (2020). Stable randomized generalized autoregressive conditional heteroskedastic models. *Econometrics and Statistics*, *15*, 67–83. http://dx.doi.org/10.1016/j.ecosta.2018.11.002.

Schmitz, H. C., Lutz, B., Wolff, D., & Neumann, D. (2023). When machines trade on corporate disclosures: Using text analytics for investment strategies. *Decision Support Systems*, *165*, Article 113892. http://dx.doi.org/10.1016/j.dss.2022.113892.

Schuster, M., & Paliwal, K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673–2681. http://dx.doi.org/10.1109/78.650093.

Tang, Y., Song, Z., Zhu, Y., Yuan, H., Hou, M., Ji, J., Tang, C., & Li, J. (2022). A survey on machine learning models for financial time series forecasting. *Neurocomputing*, *512*, 363–380. http://dx.doi.org/10.1016/j.neucom.2022.09.003.

Thakkar, A., & Chaudhari, K. (2021). A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions. *Expert Systems with Applications*, *177*, Article 114800. http://dx.doi.org/10.1016/j.eswa.2021.114800.

Thakkar, A., & Chaudhari, K. (2021). Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions. *Information Fusion*, *65*, 95–107. http://dx.doi.org/10.1016/j.inffus.2020.08.019.

Wang, J., Cui, Q., Sun, X., & He, M. (2022). Asian stock markets closing index forecast based on secondary decomposition, multi-factor analysis and attention-based LSTM model. *Engineering Applications of Artificial Intelligence*, *113*, Article 104908. http://dx.doi.org/10.1016/j.engappai.2022.104908.

Wang, L., Ma, F., Liu, J., & Yang, L. (2020). Forecasting stock price volatility: New evidence from the GARCH-MIDAS model. *International Journal of Forecasting*, *36*(2), 684–694. http://dx.doi.org/10.1016/j.ijforecast.2019.08.005.

Wang, Y., Wang, L., Yang, F., Di, W., & Chang, Q. (2021). Advantages of direct input-to-output connections in neural networks: The Elman network for stock index forecasting. *Information Sciences*, *547*, 1066–1079. http://dx.doi.org/10.1016/j.ins.2020.09.031.

Wang, H., Wang, X., Yin, S., & Ji, H. (2022). The asymmetric contagion effect between stock market and cryptocurrency market. *Finance Research Letters*, *46*, Article 102345. http://dx.doi.org/10.1016/j.frl.2021.102345.

Wei, L. Y. (2016). A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting. *Applied Soft Computing*, *42*, 368–376. http://dx.doi.org/10.1016/j.asoc.2016.01.027.

Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications*, *112*, 258–273. http://dx.doi.org/10.1016/j.eswa.2018.06.016.

Wilder, J. W. (1978). *New concepts in technical trading systems*. Trend Research.

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

Williams, B. M. (1998). *New trading dimensions: how to profit from chaos in stocks, bonds, and commodities: vol. 72*, John Wiley & Sons.

Wu, M. E., Syu, J. H., Lin, J. C. W., & Ho, J. M. (2022). Effective fuzzy system for qualifying the characteristics of stocks by random trading. *IEEE Transactions on Fuzzy Systems*, *30*(8), 3152–3165. http://dx.doi.org/10.1109/TFUZZ.2021.3105192.

Xie, C., Rajan, D., & Chai, Q. (2021). An interpretable neural fuzzy Hammerstein-Wiener network for stock price prediction. *Information Sciences*, *577*, 324–335. http://dx.doi.org/10.1016/j.ins.2021.06.076.

Yang, Y., Guo, J., Sun, S., & Li, Y. (2021). Forecasting crude oil price with a new hybrid approach and multi-source data. *Engineering Applications of Artificial Intelligence*, *101*, Article 104217. http://dx.doi.org/10.1016/j.engappai.2021.104217.

Yi, Z., Liang, Z., Xie, T., & Li, F. (2023). Financial risk prediction in supply chain finance based on buyer transaction behavior. *Decision Support Systems*, *170*, Article 113964. http://dx.doi.org/10.1016/j.dss.2023.113964.

Zhang, C. X., Li, J., Huang, X. F., Zhang, J. S., & Huang, H. C. (2022). Forecasting stock volatility and value-at-risk based on temporal convolutional networks. *Expert Systems with Applications*, *207*, Article 117951. http://dx.doi.org/10.1016/j.eswa.2022.117951.

Zhong, C., Du, W., Xu, W., Huang, Q., Zhao, Y., & Wang, M. (2023). LSTM-ReGAT: A network-centric approach for cryptocurrency price trend prediction. *Decision Support Systems*, *169*, Article 113955. http://dx.doi.org/10.1016/j.dss.2023.113955.

Zolfaghari, M., & Gholami, S. (2021). A hybrid approach of adaptive wavelet transform, long short-term memory and ARIMA-GARCH family models for the stock index prediction. *Expert Systems with Applications*, *182*, Article 115149. http://dx.doi.org/10.1016/j.eswa.2021.115149.