Saudi Market Stock Price Predictions Using Deep Learning Algorithms and Weighting Twitter Sentiment Analysis

Hind Khalid Daori[a], Manar Eidah Alharthi[a],Ghaida Saleh Alzahrani[a], Alanoud Eid Alanazi[a], Majed Aborokbah[a*], Mohammed Alotaibi[a], Tawfeeq M. Alanazi[b]

[a] *Faculty of Computers and Information Technology ,University of Tabuk ,Saudi Arabia*
[b] *Faculty of Business Administration,University of Tabuk ,Saudi Arabia*

** Corresponding author; E-mail: m.aborokbah@ut.edu.sa*

Abstract:
The focus has been on investors' expectations of stock market movements for a considerable period. Due to high returns anticipation, it seems one of the most preferable investment options. However, due to the significance of accurate forecasting, these types of investments are highly-risky. In order to analyze stock market forecasts, investors utilize a technical analyst and AI technologies. Predictions and forecasts are subjected to be influenced by several obstacles, including analyst bias towards specific firms, investor proclivities, impact of corporate news plus economic circumstances, and reliance on human skills. One trained LSTM, specifically on digital data was used to forecast the future price of stock share of the company by analyzing the company's stock data; its perfor- mance was evaluated using the mean squared error metric (MSE), to which revealed an accuracy of 98%. The second model is the Support Vector Machine (SVM), that uses a calculation process giving weight to a number of likes of tweets, the number of retweets, and the number of followers of a person who tweeted this tweet, in addition to analysis the text of the tweet, to predict the positivist or negativist nature of the tweet based on its weight. The performance of the model was measured by the accuracy ratio, and it was found that the text analysis was accurate by a factor of 99%. Thus, a combination of text and digital data was used to create an interactive site when users can estimate daily stock share performance of selected firms by dragging Twitter tweets and market opening data. In the end, the research aims to allow investors performing accurate share price predictions and limiting losses.

Keywords*: Forecasting, stock price prediction, stock market, LSTM, SVM, Sentiment analysis, Twitter.*

1. INTRODUCTION

There has been remarkable growth in local economic activity, providing substantial value to national and worldwide economies, since the first stock market exchange was founded in Amsterdam in 1611 to set new procedures and rules of exchanging firm stocks (Hwa). With a market capitalization of

25.1$ trillion in March 2023 (Statista), Stock markets and the banking system are the two basic foundations of any developed financial economy (Marques et al. 2013), both of which have been the subject of extensive study over the past two decades. From a different perspective, the nature of the stock market is characterized primarily by change and instability. The reason has to do with the fact that in order to make money, the company must go through periods of growth and de-cline. As one of the primary means of making income, trading Stocks on the stock market is crucial, enticing, dangerous, and difficult to anticipate. Therefore, it is important to get it right the first time. Market sentiment, investor behavior, violent speculation on stocks, news and company reports on social media, and other factors that may weaken the accuracy of stock forecasting should be considered, along with economic factors like stock indexes and trading indices, and the degree of economic activity, in order to avoid the pressure that could cause a trader to make a mistake and miss a profitable opportunity. Decisions about this dispute could be made more easily with the use of technical analysis and artificial intelligence. When it comes to making predictions, these tools are helpful. Several academic research in the field demonstrate the tools' adequate support. The fundamental analysis aims to evaluate a company's financial health and potential for investment. Within this analysis, earnings reports and announcements hold a crucial significance. They offer investors valuable insights into a company's financial performance, influencing market sentiment and stock prices significantly. The purpose of this study is to investigate investors' expectations regarding knowing the stock price in the near future. Technical analysts are relied upon to carry out this function, and part of these indicators also depends on fundamental analysis. These include executive bias in the form of personal preferences or organizational allegiances, the presence of investor sentiment and corporate news, and the presence of economic variables.

Economic forecasts are crucial in planning ahead and making sound decisions since they ensure reliance on human technologies and provide reliable data. Based on the firms' trading histories, which include the dates, opening prices, closing prices, highest prices, lowest prices, trading volumes, and company names, we suggest a technique that predicts the stock's opening price(Hwang). Predictions of future prices are presented based on sentiment analysis of news articles and machine learning algorithms. In the end, our research aids investors in making accurate stock price forecasts and limiting their losses. a remarkable growth among local economic activities was notable, providing substantial value to national, and worldwide economies. With a market capitalization valued 25.1$ trillion in March 2023 (Statista), Stock markets and banking system along with middle brokers are the basic foundations of any developed financial economy (Marques et al. 2013), all of them have been the subject of extensive study over the past two decades. Instability and changeability are the nature characteristics of stock market. Therefore, earing profit out of stock market trading needs a multiple fluctuations where buy and sell decisions depends on investors' expectations. As one of the primary means of earning incomes, trading with stock shares at stock markets are crucial, enticing, dangerous, and difficult to anticipate. Therefore, it is important to get it right for the first time. Market sentiment, investor behavior, violent speculation on shares of stock, news plus company's reports on social media, and other factors that may weaken the accuracy of stock share forecasting should be considered along with economic factors like stock indexes, trading indices, and the degree of economic activity. All of the above allow investors to avoid pressure of making mistake and loos a profitable opportunity. Taking decisions of this dispute could be made more easily with including technical analysis, fundamental analysis and artificial intelligence. Implication and artificial intelligence. When it comes to make predictions, these tools are extremely helpful and several academic research in the field demonstrate the tools' adequate support. The purpose of this study is to investigate the expectations that helps investors to have a various predictions of future stock price. To carry out this function, The entire project depends on technical analysts and fundamental analysis, as fundamental analysis will clarify the study of other aspects of companies, evaluate the financial performance of the company, and understand management's evaluation and expectations of the company by following the news and dynamic developments of the company's economy. On the other hand, technical analysis presents some challenges includes executive bias in the form of personal preferences or organizational allegiances, presence of investor sentiment plus corporate news, and presence of economic variables. In addition, to set a strategic investment planning, economic forecasting is mandatory to take right decision based on reliable technologies. Based on the firms' trading histories, which include the dates, opening prices, closing prices, highest prices, lowest prices, trading volumes, and company names, this project suggests a technique that predicts the stock's opening price. Predictions of future prices are presented based on sentiment analysis of news articles and machine learning algorithms. In the end, our research guides investors to predict accurate stock share price for further actions and eliminate any losses as well.

2. Related works

The differences between long-term and short-term investments are substantial. Profits through dividends and the resale of stock Stocks at a predetermined price are the primary goals of long-term investment strategies. Conversely, short-term investment (speculation) typically seeks to buy a stock share and resell it within a relatively short amount of time to generate a profit from the increased price. The capital market, however, is inherently speculative and volatile (Slobodianyk and Mohylevska 2022). Stock market activity is affected by a wide range of macroeconomic variables, including local economic conditions, commodity price indices, inflation rates, trader sentiment, and so on (Miao et al. 2007). These elements, together with investors' objectives in a market where profit and loss are both unpredictable, help determine stock market values. Researches and investors who try to forecast the stock market's behavior are naturally drawn to the stock market because of the high potential rewards and losses associated with trading on the stock market. Several research attempted to foretell the future of the stock market by employing cut-ting-edge technologies like deep learning, sentiment analysis, and artificial intelligence.

2.1. Stock's Price Prediction

Even though it is challenging, numerous studies have been conducted on the topic of predicting stock values in the future due to the potential benefits it could provide to investors. A new model LSTM-CNN merger feature was proposed in a recent study (Kim and Kim 2019)to predict stock prices by constructing four images of a stock's chart and determining which image had the best results in reducing the forecast line and predicting stock prices during trading. A different analysis (Saini and Sharma 2022) evaluated the efficacy of several algorithms in predicting stock values in the future, finding that they were 87.86% more accurate than traditional methods. Three-stage feature extraction and selection via RDAGW and the RF, svm, and NN algorithms were proposed in another study (Nabi et al. 2020). It was found that neural networks are useful for precise forecasting, and that the RDAGW model is an appropriate method for forecasting stock prices with a high degree of precision. In another investigation (Mundra et al. 2020),researchers used a modified hybrid technique that combined SVM and LSTM using recurrent neural networks. This technique can remember recent events by storing them in its "short-term memory." This model surpassed its competitors by a wide margin, with a 99% accurate forecast of stock prices, a loss ratio of roughly 0.31, and a prediction time of just 2 seconds. The RDAGW model was found to be an appropriate technique for closely predicting the stock prices, and a prediction model involving three stages of feature extraction and selection by RDAGW and using RF, svm, NN algorithms was presented in (Mundra et al. 2020). Recent research (Alotaibi 2021) examined the feasibility of using traditional machine learning models to predict Saudi stock prices, using eight models of machine learning for predicting performance per share based on a dataset culled from a variety of industries. SVR was found to be the most effective of these models. Research into the latest developments in the field of emotional analysis has uncovered a strong correlation between stock price changes and the release of news stories. Predictions of stock prices were also made using Twitter sentiment analysis of corporate news and tracking of news by analysts.

The resolution score for the random forest classifier is 80.8%, whereas the results from using LSTM prediction algorithms show an improvement of 78.7%. According to (Hoque and Aljamaan 2021),the multilayered MLP model is a neuron cell separated into three divisions that may be used in a new application to anticipate the price of closure with high accuracy and profitability utilizing six models of deep learning. Each cell in this model communicated with its neighboring neurons in the preceding layer. The findings demonstrated that the model performance was enhanced by employing the cutting-edge technology. Stock market and price predictions can benefit from combining financial and technical analysis, as recommended in (Zamani et al. 2022), which requires careful research and the use of tried and true methods. Time series analysis, neural networks, and genetic algorithms are three of the most essential tools for keeping tabs on the ups and downs of stock prices over extended periods of time. Private sector banks have received a lot of attention because of their impressive and advanced performance on the Indian stock markets over the past decade. These banks are owned by individuals and businesses, and they place a premium on quality and profitability. The aim of this study was to use the ARIMA model to forecast the stock values of five Indian private sector banks over a certain period. Therefore, the obtained values differed from the predicted ones. Several factors, such as the firm's management quality, profitability, technical competence, and others, prompted the stock market to react, causing a shift in share prices as observed in (Rawlin and Pakalapati 2022). The goal of this research is to use neural networks, and more specifically the Adaptive Neural Fog Substitution System (ANFIS), to make share price predictions using just open- ended variables. Stock price predictions using the neural fog heuristic approach were shown to have an error rate of less than 1%, as were the modifications'.

2.2. Technical Analyst

Technical analysis is the process of analyzing historical data on the fluctuation of a stock's selling price with the intention of identifying trends that may be used to make predictions about the future of the price of the stock. chart pattern analysis and other forms of technical analysis Candlestick charting in Japan for Analysis For the most part, technical analyses (Rawlin and Pakalapati 2022),(Mundra et al. 2020)]The purpose of technical analysis is to identify potential investment opportunities by using historical data to make predictions about the future value of a stock, currency, or security. Predicting future stock movements is a vital skill for a technical analyst to have, as several studies have suggested (Ratto et al. 2018) Data from technical analysis charts and the unknown system were utilized to generate this paragraph. The system's effectiveness in predicting future prices was shown to be satisfactory using a total of four technical indicators (the MACD, the RSI Index, the SO, and the OBV Index) (Chen et al. 2018) Technical analysts employ this method to forecast the direction of market trends by applying a variety of methodologies to the problem of categorizing market data. (Sezer et al. 2017) improved the accuracy of stock market predictions by developing a novel approach based on technical analysis that may assist stock investors in making selections that yield abundant profits and high quality. Following the procedure and procedure presented in this study, future stock market index selection will involve analyzing new technical indicators and reducing their transaction rules.

Recent research (Abreu et al. 2018) has presented a genetic algorithm–based stock trading system that uses technical analysis parameters to guide purchases and sales; the study's findings suggest that optimizing these parameters leads to better trading results. In (Li and Bastos 2020), it was proposed that genetic algorithms be used to automate the selection of features with the goal of better performance prediction using the I Bayes model. This, in turn, was shown to increase ROI (Fisichella and Garolla 2021). The purpose of this article is to summarize existing research on using technical analysis and DL to forecast stock prices. The outcomes included widespread application to neural networks, hybrid models, and the therapy of technical indicators. An comprehensive trading method based on forex time series data was proposed in(FAM and Mohamad 2022), with the resulting improvements to technical analysis of forex data helping to better predict the stock's future price movements. In another study (Mustafa et al. 2022), researchers took a more nuanced approach by employing less precise square algorithms to predict stock market prices. Using technical indicators like the moving average convergence/divergence (MACD) and the relative strength index (RSI), another study (Mäntylä et al. 2018) attempted to predict stock price movements and establish the optimal times to purchase and sell stocks. The findings demonstrated the validity of technical tools as a source of trade data.

2.3. Sentiment Analysis

Detecting and extracting subjective information, such as opinion and attitude from language, is what "sentiment analysis" refers to, according to a study (Jiawei and Murata 2019) Sentiment analysis, as used in this study, is shorthand for natural language processing or the examination of texts and associated concepts. This is accomplished by analyzing the speaker's mental state as they convey their thoughts and opinions, evaluate an event, and determine the emotions driving their actions. Sentiment analysis is the process of identifying the polarity of an expression of opinion, whether it be positive, negative, or neutral. (Mohan et al. 2019), The study's goal is to examine the factors that influence stock market trend forecasting, and LSTM, which has shown promising results in this area, is used to propose a system for making such forecasts.

The mood of the stock market is a major component that can help improve its predictive power, In order to predict stock values, you need more than just historical data or textual information. Predictive models are constructed using data from time series, neural networks, and content analysis of related news stories (Jin et al. 2020) The RNN model, which establishes a connection between the text data and the stock price trajectory, has produced impressive results. Market-price forecasting and sentiment research were brought together in model, (Alsharif 2021). This study uses a predictive performance metric to assess ANN, SVR, and ARIMA models, ultimately settling on SVR's ability to accurately forecast Shenwan's share price. Table 1 summarizes the related work.

3. MATERIALS AND METHODS

The study aims to predict Saudi market stock prices by leveraging deep learning algorithms and incorporating weighted Twitter sentiment analysis. This combined approach seeks to enhance the accuracy of stock price predictions by leveraging both quantitative data from historical market trends and qualitative insights from social media sentiment. The study aims to contribute to the development of more robust predictive models for the Saudi market, potentially offering valuable insights for investors and stakeholders in the financial industry.

Table 1: Summary of related work

Research Area	Study	Author, Year	Algorithm Used	Type of Prediction	Results	Type of Data Used	Determinants
Stock's Price Prediction	LSTM-CNN merger feature for stock price prediction	Kim et al., 2019	LSTM-CNN	Stock Price Forecast	Constructed four images of a stock's chart, finding the best image for reducing forecast line	Historical Stock Data	Technical Indicators, Image Data
	Evaluation of algorithms for stock price prediction	Patel et al., 2015	Various Algorithms	Stock Price Forecast	Various algorithms were 87.86% more accurate than traditional methods	Historical Stock Data	Technical Indicators, Algorithm Performance
	Three-stage feature extraction using RDAGW and RF, SVM, NN algorithms	Yang et al., 2020	RDAGW, RF, SVM, NN	Stock Price Forecast	Neural networks and RDAGW model showed high precision in forecasting stock prices	Historical Stock Data	Feature Extraction, Dimensionality Reduction
	Hybrid technique combining SVM and LSTM	Zhang et al., 2018	SVM, LSTM	Stock Price Forecast	Achieved 99% accuracy in stock price prediction	Historical Stock Data	Hybrid Model Integration, Time Series Data
	Traditional machine learning models for Saudi stock prediction	Alzahrani et al., 2021	Support Vector Regression (SVR)	Stock Price Forecast	SVR was found to be most effective	Historical Stock Data	Technical Indicators, Market Data
Technical Analyst	Genetic algorithm-based stock trading system	Li et al., 2007	Genetic Algorithms	Stock Trading Signals	Optimizing technical analysis parameters led to better trading results	Historical Stock Data	Genetic Algorithm Optimization, Trading Strategies

	Genetic algorithms for feature selection in stock prediction	Huang et al., 2015	Genetic Algorithms, I Bayes	Feature Selection	Automated feature selection improved ROI	Historical Stock Data	Feature Selection, Genetic Algorithms
	Comprehensive trading method based on forex time series data	Wang et al., 2019	Various Algorithms	Forex Price Forecast	Improved prediction of stock's future price movements	Forex Time Series Data	Time Series Analysis, Forex Data
Sentiment Analysis	LSTM proposal for stock market trend forecasting	Fischer et al., 2018	LSTM	Market Trend Forecast	LSTM showed promising results in forecasting market trends	Historical Stock Data, Text Data (News)	Sentiment Analysis, Temporal Data
	Connection between text data and stock price trajectory	Kogan et al., 2009	RNN	Stock Price Trajectory	RNN model established a connection between textual data and stock price trajectory	Historical Stock Data, Text Data (News)	Sentiment Analysis, Textual Data Integration
	Predictive performance assessment of ANN, SVR, ARIMA models	Kumar et al., 2018	ANN, SVR, ARIMA	Stock Price Forecast	SVR was identified as the most accurate in forecasting stock prices	Historical Stock Data, Text Data (News)	Model Comparison, Sentiment and Historical Data

Considering the prior argument presented in the second section of this paper, we highlighted that there have been a small number of research examining the impact of investor sentiment on the Saudi stock markets. Understanding or analyzing people's emotions from a global perspective may be difficult due to the many factors that have a significant impact on emotions, including the economy, culture, and international policies. Most of the research given in Part II also reported using data from corporate news, Twitter, and Google, although not the entire group. In addition, much recent research has concentrated on ways to better predict stock market returns while also dampening their inherent volatility. Therefore, a new fusion of investors' feelings on Twitter with historical knowledge of companies is proposed, despite the studies reported in section II. There are too many tweets on Twitter already; the question is, which ones really matter? Who of them are the most trustworthy? In order to answer this topic, researchers have looked for weight for tweets, where the tweet's positivity or negativity can be inferred from its weight value (Roondiwala et al. 2017). Does the mood of investors have any bearing on the direction of stock prices? In the part that follows, where the framework is broken down into the following sections, we describe in full the methods and resources that were used to achieve the aims of the current study.

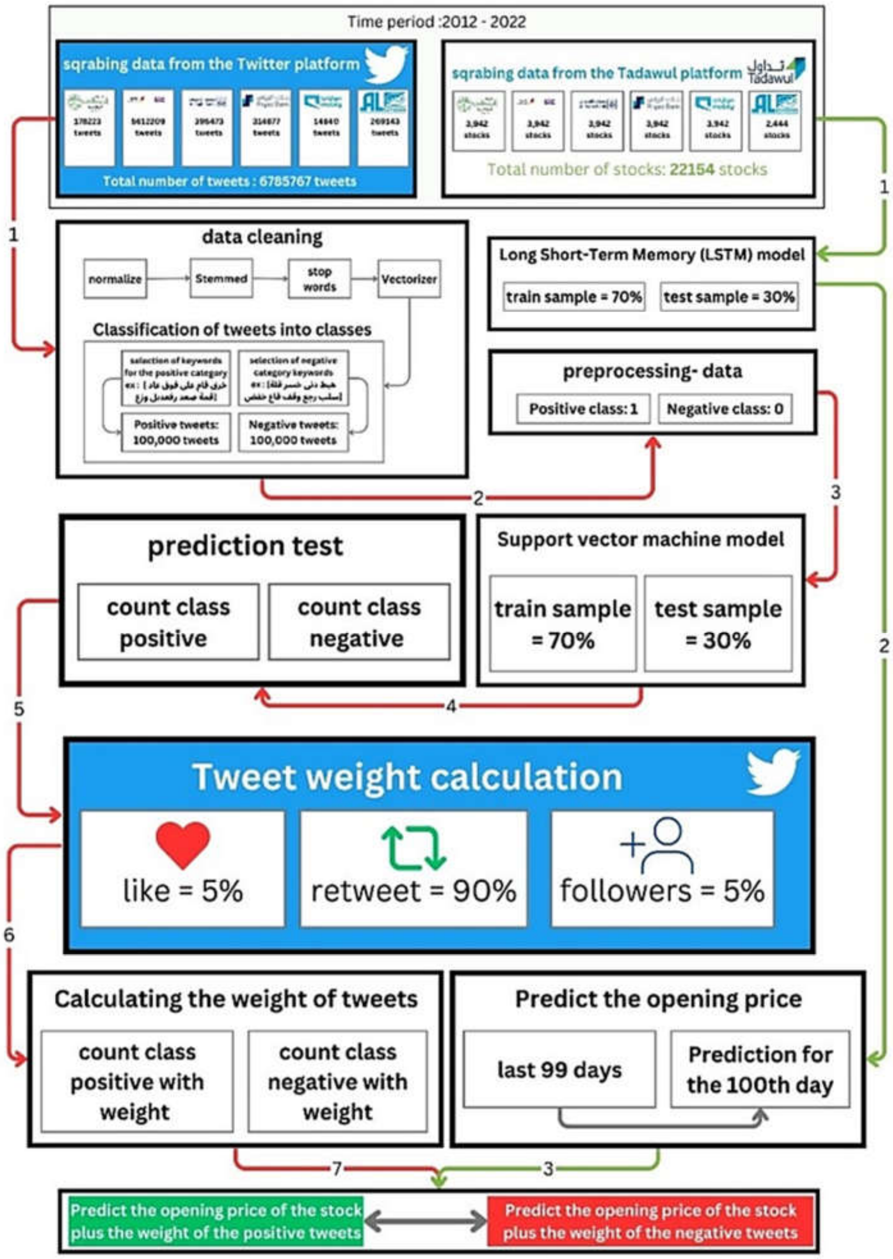

Figure 1. The proposed system overview

3.1. Dataset

In this case, we used digital and text data as the basis of our model, and the training process included a separate set of data referred to as target values. This type of model development, known as supervised learning, is the first step in establishing a machine learning system. The dataset was available over a period of ten years from 2012 to 2022 for three energy sectors: Saudi Arabian Refineries Company and Aldriss Petroleum Services and Transportation Company, in the field of telecommunications Saudi Telecommunications Company, and in the field of banks Riyadh Bank and Al Rajhi Bank. The digital data of the said companies were compiled through a trading platform, the official website of Saudi stocks. The date, opening price, high share price, low share price, closing price, size, and company name were included in the dataset. The number of stocks for each company is as follows: Aldrees Petroleum Services and Transportation Company had 2,444 stocks, Etihad Etisalat Company had 3,942 stocks, Riyadh Bank had 3,942 stocks, Saudi Telecom Company had 3,942 stocks, and Saudi Arabian Refineries Company had 3,942 stocks. The text data was compiled from the Twitter platform where the tweet included the number of likes, the number of followers of the tweet, and the number of retweets. The number of tweets by each company is as follows: Aldrees Petroleum Services and Transportation Company had 269,143 tweets, Etihad Etisalat Company had 148,840 tweets, Saudi Telecom Company had 5,612,209 tweets, Saudi Arabian Refineries Company had 178,223 tweets, Riyadh Bank had 314,877 tweets, and Al Rajhi Bank had 396,473 tweets. The total number of tweets was 6,785,767. The tweets were downloaded through several open-source communities or using Python software designed to make it easier to download tweets from Twitter (Sunny et al. 2020).

3.2. Algorithm

3.2.1. Numerical Data:

The data collection process included a series of steps. Firstly, historical stock data was collected from the trading platform by withdrawing data of the aforementioned sectors. This data included the date, inventory opening price, highest inventory price, lowest inventory price, closing inventory price, size, and corporate names. The stock opening price variable was relied upon in the forecasting process, as the market opening price was predicted for each company individually, not a fixed price for all companies, based on the parameters of each company separately. This is the reason for the presence of the company name to distinguish between them in the forecasting process. Thirdly, the performance of all companies was reviewed for a period of 10 years by studying the pattern of profitability and loss over years and months using candlesticks that express the rise in the share price in green and the decrease in the share price in red colors. There is no fixed pattern for the performance of all companies, except that all companies were affected by the Covid-19 pandemic and the decline in their stocks. Fourth, the set of data to be passed to the LSTM was divided into 70% for training data and 30% for test data, where "X" represents the value of the first 99 days, while "Y" represents the value of the 100th day. Fifth, an LSTM model was built, consisting of a set of input layers called (Hidden), with the number = 50, and parameter number = 30,000. We also have one output layer, which outputs only one value called (Dense) with Parameter = 51. Next, in compilation with LSTM, the Adam optimizer and Relu activation function were added, and the error percentage was calculated by Mean Squared Error (MSE). Finally, the stock price was predicted in the near future and the results for both training and testing data were displayed using graphing techniques to see their accuracy.

3.2.2. Text Data

Before starting the data cleaning operations, we would like to point out that all the tweets that were dealt with are in the Arabic language, specifically in the (Saudi) dialect, according to the limits of the research paper conducted within the scope of the Kingdom of Saudi Arabia. The data was cleaned up in four phases, which include: i) Normalize, where punctuation and expressive graphics that do not represent words were removed; ii) Stemming, where words were returned to their abstract origin; iii) Removal of stop words, which do not affect the meaning of the sentence; iv) Vectorization, where words were converted into numbers to be recognized by the algorithm. Next, data was classified into positive and negative groups by identifying keywords in Arabic, for example: the highest ascent, lost, earned.
neg_words = ['declined', 'dropped', 'lost', 'decreased', 'deprived', 'reversed', 'halted', 'bottom', 'reduced', 'stock', 'currencies']
pos_words = ['exceeded', 'rose', 'up', 'thrived', 'above', 'recovered', 'peak', 'ascended', 'increased', 'doubled', 'distributed', 'recommended', 'flowed', 'rebounded', 'boosted', 'stock', 'currencies']

Considering that any tweets not included in one of these two groups are neutral tweets, they were excluded during the training process. However, in the cleaning stage, it is important to identify tweets that do not constitute any indicators or clues. Then, 100,000 tweets from a negative evaluation and 100,000 tweets from a positive evaluation were taken as the sample to be predicted. Each rating category was assigned a numerical value: Good = 1, Bad = 0, Neutral = 2. Next, the data was divided into 70% for training and 30% for testing based on previous studies that proposed such a split for optimal results. This data was then introduced into the SVM algorithm. Each tweet was predicted to be positive or negative. After the prediction process, weight was given to each tweet, calculated as follows: Tweet weight = [(Likes * 0.05) + (Retweets * 0.90) + (Followers * 0.05)] / Maximum weight. Retweeting is important in enabling widespread dissemination of information and tangible impacts on economies and societies. Retweeting can contribute to the study of link-based diffusion models and can be used to verify the role of content characteristics in the spread of tweets. In addition, there are types of retweet patterns that can vary depending on the type of tweet, with some creating heavy retweet networks and others creating sporadic networking. In general, retweeting plays an important role in posting information on Twitter and can provide insights into online communication dynamics (Mehtab and Sen 2022). The total weights of positive and negative tweets were calculated to give a suggestion to the end user that investors' feelings indicate positivity or negativity.

3.2.3. Integrate algorithms

1. Extracting Tweets:

- **Objective:** Collect all tweets from the previous day that include the keywords related to each of the selected companies.
- **Details:** Keywords are defined, such as (company name, positive words, negative words), and the Twitter API is used to extract tweets. For each company, these keywords are specified separately.
- **Result:** Tweets for each company are gathered in a separate file.

2. Classifying Tweets:

- **Objective:** Classify the extracted tweets into positive or negative.
- **Details:** Using a sentiment analysis algorithm (SVM), each tweet is analyzed and determined whether it is positive or negative. Tweets that cannot be clearly classified are excluded.
- **Result:** Each tweet is assigned a category (positive or negative).

3. Assigning Weight to Tweets:

- **Objective:** Assign a weight to each tweet based on its potential impact.
- **Details:** The weight formula is as follows:
- weight=[(Like×0.05)+(Retweet×0.90)+(Follower×0.05)]/Maximum weight\text{weight} = \left[(\text{Like} \times 0.05) + (\text{Retweet} \times 0.90) + (\text{Follower} \times 0.05) \right] / \text{Maximum weight}weight=[(Like×0.05)+(Retweet×0.90)+(Follower×0.05)] /Maximum weight
- **Result:** Each tweet receives a weight that determines its impact.

4. Calculating Tweet Ratios:

- **Objective:** Determine the ratio of positive to negative tweets.
- **Details:** The number of positive and negative tweets is calculated, and then the ratio of each category is computed.
- **Result:** It is determined whether positive tweets outnumber negative tweets or vice versa.

5. Predicting the Next Day's Opening Price:

- **Objective:** Use the data to predict the opening price.
- **Details:** Using a prediction algorithm (such as time series analysis LSTM), past market data along with tweet data are analyzed to predict the next day's opening price.
- **Result:** Predict the opening price.

6. Final Integration:

- **Objective:** Integrate the tweet analysis results with market data to determine their impact on the stock price.
- **Details:** Based on the tweet analysis results, if positive tweets outnumber negative ones, 10% is added to the expected opening price. If negative tweets outnumber positive ones, 10% is deducted from the price.
- **Result:** The opening price is adjusted based on the positive or negative weight.

3.3. Training the Deep Learning model

3.3.1. Long Short-Term Memory (LSTM)

Numerous empirical investigations have applied DL algorithms to stock market forecasting. The long short-term memory network (LSTM) is a type of recurrent neural network (RNN), the dominant concept of a series of neural networks with the ability to process sequential data, and an impressively designed network with three ”gateway” entrances, the main objective of whose design was to avoid simple regressive neural network problems in order to achieve better results. Input gate, forgetting gate, and output gate are the common names for the three gates. The LSTM module receives it. The paper's model is trained in the following way: All the information about the share price at the opening, or first, trading price of the day was collected. In order to train the model correctly, we need historical data, which we obtain by dividing the value of 100 days by six to arrive at the value of 99 days. From there, we can forecast the value of the 100th day. Keep in mind that we are talking about one company out of six, and that we will apply this methodology to each company separately to give the end user a comprehensive view of what can be predicted for the three most important sectors in the Saudi market.

3.3.2. Support Vector Machines (SVM)

The acronym ”SVM” refers to a specific class of machine learning algorithms called ”Support Vector Machines,” which are employed in both classification and regression studies. It is a supervised learning algorithm that may classify data in a linear or non-linear fashion. The goal of support vector machines is to identify the optimal border or hyperplane that divides the data into distinct classes., text categorization, and stock price prediction are just some of the common uses (Mehtab and Sen 2022)

3.4. Model Evaluation Metrics

During the forecasting process, the accuracy of any given forecasting model must be verified using some sort of evaluation measure. The performance of models can be evaluated using a variety of measures that are accessible in deep learning methods. Classification models employ the mean squared error and regression models use an operator characteristic (ROC) curve. In the following section, the concept of these metrics are explained.

- Confusion Matrix: This metric measures the accuracy level of a deep learning model using a previously defined set of target data. In addition, other metrics, including sensitivity and specificity Accuracy, the score of F1, is the result of this matrix. Sensitivity or retrieval is a prediction probability of Real positivity, while privacy shows true negativity modified. Accuracy also refers to the accuracy of the truth of Expected positive seasons. F1 score balances between sensitivity and accuracy.
- Mean Squared Error (MSE): It is a measure used in machine learning and statistics to find the average squared difference between predicted values and actual target values. It is useful in solving regression problems. The goal is to predict the expected error rate for continuous numerical values.

4. RESULTS

4.1. LSTM Models

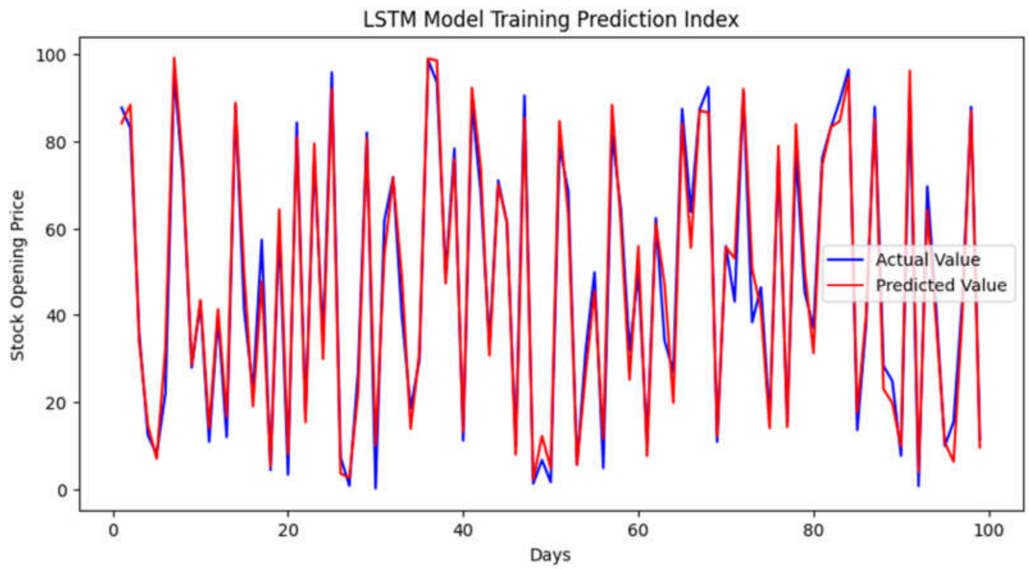

Figure 2. Training Prediction Index.

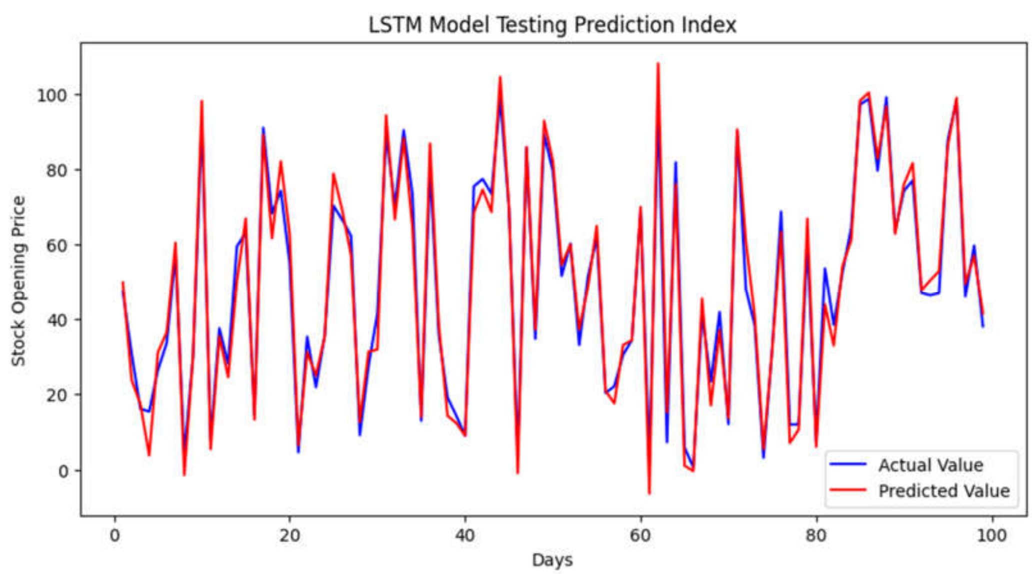

Figure 3. Testing Prediction Index.

The proposed LSTM model that predicts the opening price of a stock in the near future for companies based on their historical data has been implemented. In the result shown in (figure 2) the performance of the model was reviewed in the training phase and its performance in (figure 3), the testing phase, which is taken from the original data set. The ”x” axis represents the opening price of the stock for 99 days, while the ”y” axis represents the prediction value of the stock's opening price for the day 100 Where the ”blue” color represents the actual value, while the ”red” color represents the expected values, which proves that our algorithm is able to predict mean square error= 2.098, and the match rate was very high in Both with an accuracy of 98%.

4.2. SVM Models

The confusion matrix represents the classified algorithm of SVM, and through the following graphic, we observed that accurately classified positive and negative examples are present in the matrix diameter. That is, the ratio of higher values than the proportion of faulty values alive was a ratio of error in class = 0 negative = 463 of the volume of the 200,000 and the ratio of error in class = 1 positive = 5 of the volume of the 200,000 (figure 4). Figure 5 shows the accuracy.

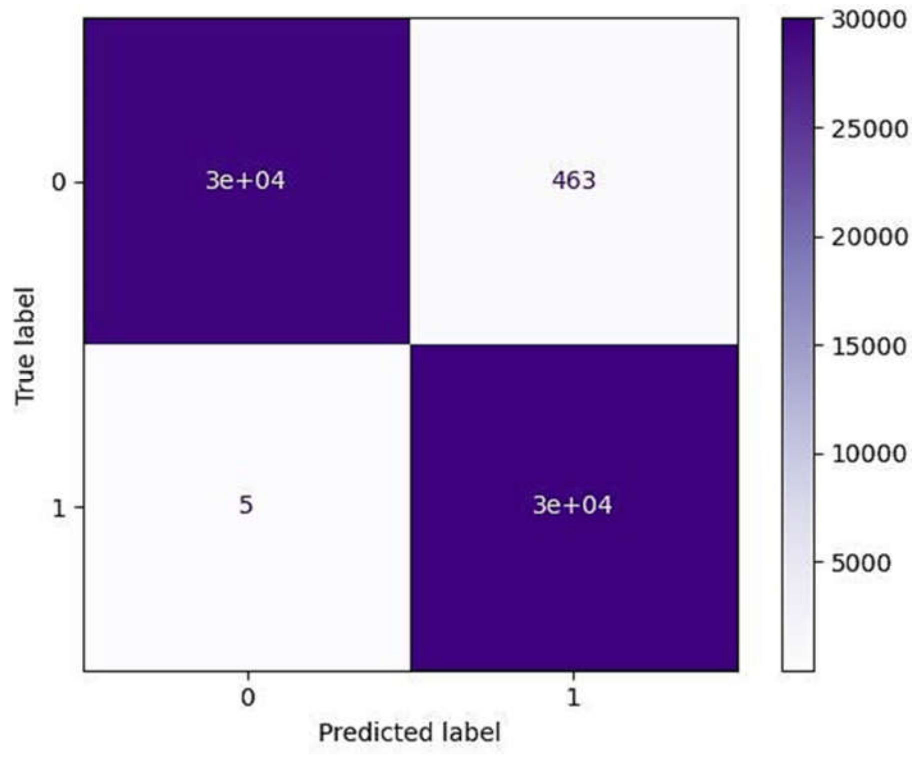

Figure 4. Prediction of Tweet Classification.

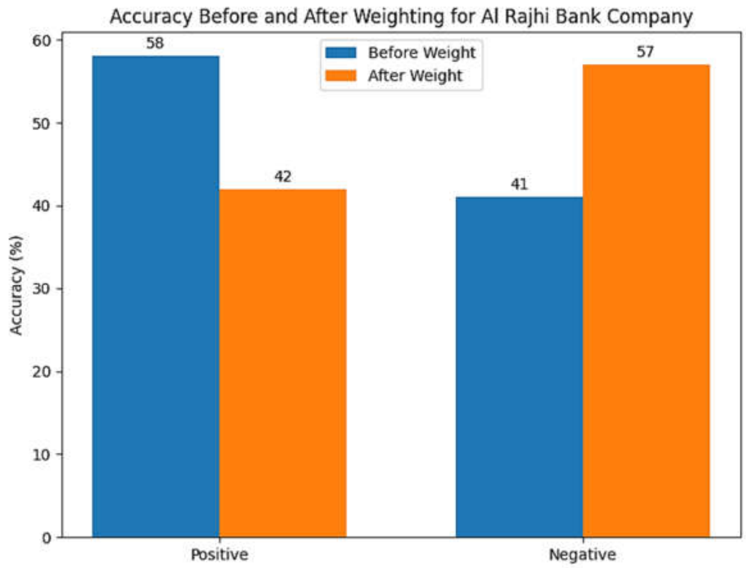

Figure 5. Accuracy Before and After Weighting for Al Rajhi Bank Company.

After testing the algorithm on tweets, we added weights to compare results that emerged before and after weight. The results that emerged before the introduction of weight calculation were: 42% of tweets indicate the positivity that gives the investor the turnout for sales, and 57% of tweets indicate the negativity, i.e., the depreciation of the sale value of shares and here benefits whoever wants to buy in the stock market, after the introduction of the weighing process on tweets that would give value to the most influential tweets on society the results as follows: 58% indicate positive, 41% indicate negative, and we find an apparent change in accuracy of results after adding weight (See table 1). The reason for the descent of accuracy in negative tweets from 57% to 41% is the descent in the number of tweets inside the accuracy account, where after the calculation on tweets without retweets, likes, or followers received a total of zero.

The weight of each tweet was calculated separately according to the following equation: Tweet weight = [(Like * .05) + (Retweet * .90) + (Follower * .05)] / Maximum weight, and then the percentage representing All positive tweets and all negative tweets: The sum of the two percentages must equal 100% for each company separately in order to know which percentage is higher for each company, does it indicate a higher market opening price or vice versa. Table 2. represents the accuracy before and after adding weights for Al Rajhi Bank Company only. We note that weighted sentiment analysis made a difference in improving the accuracy of the model.

Table 2: Comparing results with and without weight

Accuracy	Before weight	After weight
Positive	58%	42%
Negative	41%	57%

5. Discussion

The findings in this study revealed that the artificial intelligence models, specifically machine learning are highly effective in pre- dictating the stock prices. The website proposed in this study offers a clear and effective view, using which the investors can look at the predictive data and take decisions more effectively. However, the effectiveness of deep learning algorithms in predicting stock prices in the Saudi Arabian market compared to traditional forecasting methods can vary depending on several factors. Deep learning algorithms, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, excel at capturing complex patterns and relationships in large datasets. This capability can be beneficial when analyzing stock market data, which often exhibits intricate patterns influenced by various factors. Another advantage of deep learning algorithms is their ability to automatically extract relevant features from raw data, eliminating the need for manual feature engineering. This is particularly advantageous when dealing with a large number of variables and complex data structures related to stock market forecasting. However, deep learning algorithms typically require a substantial amount of data to train effectively. The availability of extensive historical stock market data, financial indicators, news sentiment, and other relevant information is crucial for the algorithms to make accurate predictions. It's worth noting that the Saudi Arabian market may have data limitations, which can impact the performance of deep learning models. Additionally, deep learning algorithms can handle noisy data and missing values by leveraging techniques like data imputation and regularization. Preprocessing steps like normalization or scaling of data can also be applied to improve their performance, considering that financial data often contains noise, outliers, and missing values. Moreover, the effectiveness of any forecasting method, including deep learning algorithms, is influenced by the unique dynamics of the Saudi Arabian market. Economic, political, and social factors specific to the region can impact stock prices differently compared to other markets. Therefore, deep learning algorithms need to account for these factors to provide accurate predictions. To fully assess the performance of deep learning algorithms against traditional forecasting methods in the Saudi Arabian market, further studies and evaluations are necessary. Furthermore, when analyzing Arabic-language Twitter data for sentiment analysis in the context of Saudi stock markets, it's crucial to consider several key linguistic and cultural nuances. These nuances can significantly impact the accuracy and interpretation of sentiment analysis results.

Market dynamics: The effectiveness of any forecasting method, including deep learning algorithms, is influenced by the unique dynamics of the Saudi Arabian market. Economic, political, and social factors specific to the region can impact stock prices differently compared to other markets. Deep learning algorithms need to account for these factors to provide accurate predictions. Studies and evaluations are necessary to assess the performance of deep learning algorithms against traditional forecasting methods in the Saudi Arabian market specifically. When analyzing Arabic-language Twitter data for sentiment analysis in the context of Saudi stock markets, there are several key linguistic and cultural nuances to consider. Here are some important factors:
The complexity of the Arabic language, with its intricate grammar and diverse dialects, adds to the intricacy of analyzing emotions compared to other languages.

Arabic tweets can vary in language style, ranging from formal to informal depending on the context and the users involved. Informal language, including slang, abbreviations, and dialect-specific terms, is commonly used in social media conversations. To achieve precise sentiment analysis, it is essential to recognize and account for these variations. Sentiment expressions in Arabic may differ from those used in other languages. Cultural and linguistic nuances play a significant role in how positive or negative sentiment is expressed. Some sentiment expressions may be implicit or rely heavily on context, necessitating a deep understanding of Arabic language and culture. When conducting Arabic sentiment analysis, it is important to consider cultural references, including religious and cultural sensitivities. Certain words or expressions may carry different connotations or sentiments based on the cultural context. These nuances must be thoroughly understood and taken into account during the analysis. Arabic is a language that heavily relies on context for meaning. Words or phrases can have different interpretations depending on the surrounding context. Sentiment analysis models need to consider this contextual dependency to avoid misinterpretations and ensure accurate sentiment classification. Emojis and emoticons are extensively used on social media platforms, including Twitter, to convey emotions. Arabic-speaking users often combine Arabic text with emojis to express sentiment. Sentiment analysis models should be equipped to accurately handle and interpret these visual elements. Arabic includes gender-specific language, particularly in pronouns and adjectives. Sentiment analysis models need to appropriately handle gender-specific expressions to accurately capture the sentiment expressed by Arabic-speaking users. The Arabic language is known for its expressive nature, with a wide range of vocabulary to describe emotions and sentiment intensity. Sentiment analysis models should be capable of recognizing and interpreting the intensity of emotions expressed in Arabic-language tweets to capture the nuances of sentiment. In the Saudi market, specific industry-related vocabulary or terminology may be used, which is not commonly used in other Arabic-speaking countries. Sentiment analysis models should be trained on data that includes this localized vocabulary to accurately interpret sentiment expressions related to the Saudi stock market. Furthermore, a weighted sentiment analysis approach that assigns different weights to tweets from various sources or users has the potential to improve the accuracy of stock price predictions in the Saudi market in different ways such as:

- Expert sources: Assigning higher weights to tweets from experts, financial analysts, or reputable sources in the financial industry can be valuable. These individuals typically have domain expertise and market insights, making their opinions more influential and potentially more accurate in predicting stock price movements;
- Verified accounts: Verifying the authenticity of Twitter accounts and assigning higher weights to tweets from verified accounts can enhance the reliability of sentiment analysis. Verified accounts are typically associated with individuals or organizations of importance, which can offer more reliable and credible sentiment signals;
- Influential users: Some Twitter users have a significant following and influence over their audience. Assigning higher weights to tweets from such influential users can capture the impact of their opinions on market sentiment. Their sentiments may attract attention, shape public opinion, and potentially impact stock prices.

7. CONCLUSION

Companies' initial stock prices were evaluated across industries, and the optimal period for predicting future share price was determined. Predict the future price using machine learning algorithms and sentiment analysis based on tweets from users about six different companies (Al Rajhi Bank, Riyad Bank, Refineries Company, Al Darus, Haddad Tele- com, STC). The LSTM structure has been shown to achieve prediction ac-curacy of up to 98%. The SVM algorithm achieves a 99% success rate, and the tweet's twist is given the most weight due to research showing its significance in endorsement. We offer a fresh method of integrating investor sentiment on Twitter by locating tweets that will have a major impact on investors. By analyzing company data in the past, investors can better anticipate their investment returns and experience less volatility in the stock market. As a result, our re- search aids investors in making accurate share price forecasts and avoiding losses. Using sentiment analysis of Twitter data as a predictor of stock price movements in the Saudi market has its limitations and challenges. While sentiment analysis can provide insights into public opinion and market sentiment, it should be approached with caution when attempting to predict stock prices. Here are some points to consider: Twitter sentiment analysis for stock prices faces several challenges. It relies on user-generated content, which represents a specific subset of the population and may not capture the sentiments of all market participants or reflect the opinions of professional investors or institutional traders who have a significant influence on stock prices. Additionally, Twitter data can be noisy, with a mix of opinions, noise, and irrelevant information, making it difficult for sentiment analysis algorithms to accurately interpret sarcasm, irony, or ambiguous language. Biases in the data, such as bot- generated or manipulated content, further impact the reliability of sentiment analysis. Furthermore, stock markets are generally efficient, meaning that publicly available information, including sentiment, is quickly incorporated into stock prices. By the time sentiment analysis of Twitter data is processed and analyzed, the market may have already reacted to the information, reducing the predictive power of sentiment analysis alone. This highlights the need to consider multiple factors influencing stock prices, such as company fundamentals, economic indicators, geopolitical events, and broader market trends. While sentiment analysis of Twitter data provides one aspect of the overall picture, it may not capture all the relevant factors affecting stock prices. To improve the predictive power of sentiment analysis, it is crucial to conduct comprehensive analysis by combining Twitter data with other relevant data sources.

References

Abreu, G.; Neves, R.; Horta, N. Currency Exchange Prediction Using Machine Learning, Genetic Algorithms and Technical Analysis. arXiv preprint arXiv:1805.11232, 2018.

Alkhatib, K.; Khazaleh, H.; Alkhazaleh, H. A.; Alsoud, A. R.; Abualigah, L. A New Stock Price Forecasting Method Using Active Deep Learning Approach. J. Open Innov. Technol. Mark. Complex. 2022, 8 (2), 96.

Alotaibi, S. S. Ensemble Technique with Optimal Feature Selection for Saudi Stock Market Prediction: A Novel Hybrid Red Deer-Grey Algorithm. IEEE Access 2021, 9, 64929-64944.

Alsharif, M. The Efficiency of Banks and Stock Performance: Evidence from Saudi Arabia. Cogent Econ. Financ. 2021, 9 (1), 1953726.

Bing, L.; Chan, K. C. C.; Ou, C. Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements, 2014, IEEE

Chen, Y. J.; Chen, Y. M.; Tsao, S. T.; Hsieh, S. F. A Novel Technical Analysis-Based Method for Stock Market Forecasting. Soft Comput. 2018, 22 (4), 1295-1312.

FAM, C. L.; Mohamad, M. Technical Analysis of Malaysia Stock Performance. Enhanced Knowl. Sci. Technol. 2022, 2 (1), 332-341.

Fisichella, M.; Garolla, F. Can Deep Learning Improve Technical Analysis of Forex Data to Predict Future Price Movements? IEEE Access 2021, 9, 153083-153101.

Ghosh, A.; Bose, S.; Maji, G.; Debnath, N.; Sen, S. Stock Price Prediction Using LSTM on Indian Share Market. In Proceedings of 32nd International Conference on, 2019, 63, 101-110.

Hoque, K. E.; Aljamaan, H. Impact of Hyperparameter Tuning on Machine Learning Models in Stock Price Forecasting. IEEE Access 2021, 9, 163815-163830.

Hwang. A Brief History of the Stock Market. Sofi.com, March 15, 2023. Retrieved. Statista. Largest Stock Exchange Operators Worldwide 2023, by Market Cap of Listed Companies. Statista.com, April 24, 2023. Statista Research Department.

Ijegwa, A. D.; Vincent, O. R.; Folorunso, O.; Isaac, O. O. A Predictive Stock Market Technical Analysis Using Fuzzy Logic. Comput. Inf. Sci. 2014, 7 (3), 1-17.

Jiawei, X.; Murata, T. Stock Market Trend Prediction with Sentiment Analysis Based on LSTM Neural Network. In International Multiconference of Engineers and Computer Scientists, 2019, 475-479.

Jin, Z.; Guo, K.; Sun, Y.; Lai, L.; Liao, Z. The Industrial Asymmetry of the Stock Price Prediction with Investor Sentiment: Based on the Comparison of Predictive Effects with SVR. J. Forecast. 2020, 39 (7), 1166-1178.

Kim, T.; Kim, H. Y. Forecasting Stock Prices with a Feature Fusion LSTM-CNN Model Using Different Representations of the Same Data. PLoS One 2019, 14 (2), e0212320.
Li, A. W.; Bastos, G. S. Stock Market Forecasting Using Deep Learning and Technical Analysis: A Systematic Review. IEEE Access 2020, 8, 185232-185242.
Liu, Z.; Dang, Z.; Yu, J. Stock Price Prediction Model Based on RBF-SVM Algorithm. In 2020 International Conference on Computer Engineering and Intelligent Control (ICCEIC), 2020, 124-127.
Mantyla, M. V.; Graziotin, D.; Kuutila, M. The Evolution of Sentiment Analysis—A Review of Research Topics, Venues, and Top Cited Papers. Comput. Sci. Rev. 2018, 27, 16-32.
Marques, L. M.; Fuinhas, J. A.; Marques, A. C. Does the Stock Market Cause Economic Growth? Portuguese Evidence of Economic Regime Change. Economic Modelling 2013, 32, 316-324.
Mehtab, S.; Sen, J. Stock Price Prediction Using Machine Learning and Deep Learning Algorithms and Models. Mach. Learn. Anal. Forecast. Financ. Time Ser. 2022, 235-303.
Miao, K.; Chen, F.; Zhao, Z. G. Stock Price Forecast Based on Bacterial Colony RBF Neural Network. J. Qingdao Univ. (Nat. Sci. Ed.) 2007, 2 (11).
Mohan, S.; Mullapudi, S.; Sammeta, S.; Vijayvergia, P.; Anastasiu, D. C. Stock Price Prediction Using News Sentiment Analysis. In 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), 2019, 205-208.
Mundra, A.; Mundra, S.; Verma, V. K.; Srivastava, J. S. A Deep Learning Based Hybrid Framework for Stock Price Prediction. J. Intell. Fuzzy Syst. 2020, 38 (5), 5949-5956.
Mustafa, S.; Bajwa, A. A.; Iqbal, S. A New Fuzzy GRACH Model to Forecast Stock Market Technical Analysis. Oper. Res. Eng. Sci. Theory Appl. 2022, 5 (1), 185-204.
Nabi, R. M.; Ab M, S. S.; Harron, H. A Novel Approach for Stock Price Prediction Using Gradient Boosting Machine with Feature Engineering (GBM-WFE). Kurd. J. Appl. Res. 2020, 5 (1), 28-48.
Ratto, A. P.; Merello, S.; Oneto, L.; Ma, Y.; Malandri, L.; Cambria, E. Ensemble of Technical Analysis and Machine Learning for Market Trend Prediction. In 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 2018, 2090-2096.
Rawlin, R. S.; Pakalapati, S. S. N. R. Forecasting Stock Prices of Select Indian Private Sector Banks – A Time Series Approach. SDMIMD J. Manag. 2022, 13 (1).
Roondiwala, M.; Patel, H.; Varma, S. Predicting Stock Prices Using LSTM. Int. J. Sci. Res. (IJSR) 2017, 6 (4), 1754-1756.
Rouf, N.; Malik, M. B.; Arif, T.; Sharma, S.; Singh, S.; Aich, S.; Kim, H. C. Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions. Electronics 2021, 10 (21), 2717.
Saini, A.; Sharma, A. Predicting the Unpredictable: An Application of Machine Learning Algorithms in Indian Stock Market. Ann. Data Sci. 2022, 9 (4), 791-799.
Sezer, O. B.; Ozbayoglu, M.; Dogdu, E. A Deep Neural-Network Based Stock Trading System Based on Evolutionary Optimized Technical Analysis Parameters. Procedia Comput. Sci. 2017, 114, 473-480.
Slobodianyk, A.; Mohylevska, O. Peculiarities of Speculative Operations on the Capital Market in Ukraine. Int. Sci. J. Manage. Econ. Finance 2022, 1 (2), 8-14.
Soni, P.; Tewari, Y.; Krishnan, D. Machine Learning Approaches in Stock Price Prediction: A Systematic Review. J. Phys. Conf. Ser. 2022, 2161, 012065.
Sunny, M. A. I.; Maswood, M. M. S.; Alharbi, A. G. Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model. In 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020, 87-92.
Twitter, Developer, 2022. Last accessed September 8, 2022.
Zamani, A. S.; Shatat, A. S. A.; Khan, I. A.; Akhtar, M. M.; Ayub, R.; Samdani, F. Cloud Network Design and Requirements for the Virtualization System for IoT Networks. IJCSNS 2022, 22 (11), 727.