Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

STOCK MARKET PREDICITION USING STATISTICAL & DEEP LEARNING TECHNIQUES

MARAM GHIDAN ALQAHTANI[1] and HODA AHMED ABDELHAFEZ[2]

[1] College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, KSA
[2] College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, KSA and Faculty of Computer and Informatics, Suez Canal University, Ismailia, Egypt
Email: [1]maramalq@outlook.com , [2]hodaabdelhafez@gmail.com
Corresponding Author: Hoda Abdelhafez

Abstract

Predicting stock marketing prices has persistently a challenge due to the complexity of the stock data. Accurately predicting a stock's short-term price can increase the rate of investment and business opportunities in the stock market. This study aims to predict the closing prices of six major sectors in the Saudi stock market: Banking, Basic Materials, Real Estate Management and Development, Insurance, Energy, and Telecommunication. The dataset was historical records of the six sectors for seven years, along with two economic indicators: oil prices and inflation rates. Six models were employed for prediction: Auto-Regressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), Random Forests, Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), and gated recurrent units (GRU). The models were evaluated using four regression metrics: mean squared error (MSE), mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean squared error (RMSE). The findings revealed that GRU and Random Forests exhibit superior performance across multiple sectors, while SVR and Bi-LSTM demonstrated promising results. However, ARIMA consistently performed poorly across all sectors. The study provided valuable insights into the effectiveness of different models in predicting stock prices in the Saudi stock market. These findings could aid investors, analysts, and decision-makers in making informed investment decisions.

Keywords: *Stock Market Prediction; Machine Learning; AREMA; Deep Learning Techniques.*

1. INTRODUCTION

The stock market plays a vital role in the economy, and the primary goal of investors is to achieve high returns while minimizing potential losses [1]. Therefore, countries should focus on enhancing their stock markets, as doing so has been linked to economic growth [2]. Predicting the stock movement gains a great interest for decades from both investors and researchers because of its high value in seeking to maximize stock profit [3]. Various economic factors such as economic conditions and political events and investors' expectations influence the stock market movements [1]. Given the possibility of achieving rapid returns on investment, accurately predicting the stock market presents a viable avenue toward attaining financial independence.

Stock market prediction is the process of forecast future trends and movements of the stock market to make profitable investment decisions. This task is critical in financial markets, where investors and traders seek to make informed decisions about their investments [4]. Technical analysis of the stock market is used to predict future stock prices by examining historical data and using technical indicators to identify patterns and trends [4]. Recent studies have shown that the historical records of stock prices can be utilized to make future predictions with a high degree of accuracy [5], [6]. Consequently, there is a growing interest among researchers and investors in developing models and techniques that can provide more accurate predictions and enhance investment strategies. The models developed are categorized into distinct methodologies, including statistical models, machine learning models, deep learning models, and hybrid models [7]. The complexities inherent in the global financial markets are an endless source of attraction and an area of significant research focus in the economic domain [8]. The Saudi stock market, known as Tadawul, is one of the largest and most influential markets in the Middle East and North Africa (MENA) region, and plays a pivotal role in the global economic landscape [9]. The Tadawul All Share Index (TASI), which serves as the market's leading indicator includes several industries, each of which has its own distinct traits, tendencies, and future directions [10]. The performance of Saudi stock market significantly effects the economy and individual investors. However, accurately predicting stock market prices presents notable challenges because of various factors, which include the nonlinear nature of stock data, the impact of economic conditions on market dynamics, and the need for enhanced data quality [1], [7]. Moreover, the prediction of financial time series becomes a problem because of complex features such as volatility, changing trends and irregularities [7]. Thus, addressing these data quality issues is essential to ensure robust and accurate predictions.

On the other hand, predictive modelling in finance is a complex, multifaceted field concerned with anticipating future financial trends and outcomes based on historical data and current market conditions as well as other external factors [1]. Due to this, Traditional statistical models frequently do not reflect the complicated linkages between these variables and market pricing. With the development of modern computing capabilities, Machine learning and deep learning has emerged as a viable alternative to these traditional methods, delivering sophisticated algorithms that can learn from data and improve their predicting accuracy over time. Extract features from data and identify hidden nonlinear relationships without depending on human expertise and econometric assumptions makes deep learning techniques more attractive compared with other existing models [11] Therefore, the aim of this research study is to address the limitations of traditional methods by proposing and analyzing different machine learning and deep learning techniques for predicting Saudi stock market prices. Whereas the stock market is significantly influenced by economic and geopolitical factors such as fluctuations in oil prices, inflation rates, political and security changes. These external factors incorporated alongside historical data to enhance the accuracy of predictions. Thus, this study focused on the essential features related to the Saudi stock markets parameters, such as the indicator that gauges the overall performance of a specific financial sector within the market and the opening and closing price representing the stock price at the beginning and the end of a trading session. The highest and lowest prices reached during the trading session, the total volume and value of shares traded, and the total number of individual trades executed for a particular stock during a trading session. Furthermore, two features have been added due to their influence on the exchange: oil prices and inflation rate. These features are crucial in understanding the Saudi stock market's behavior, tracking the market's performance, and making informed investment decisions. The other factors are not covered in this study. Predicting the closing price included six major sectors in the Saudi stock market, which are banking, basic materials, Real Estate management and development, telecommunication, insurance, and energy. The implemented six prediction models were ARIMA, SVR, Random Forests, LSTM, Bi-LSTM, and GRU.

The remaining sections of this paper are organized as follows: Section 2 contains related studies of the stock market prediction. Section 3 describes research methodology including feature selection, the prediction models, and the evaluation criteria. Section 4 describes the data collection, exploration pre-processing, implementation of the models and evaluation. Section 5 focuses on the discussion of the results and finally conclusion and future work section.

2. Related Work

Many researchers have been applying statistical, machine learning, and deep learning models to enhance different aspects of stock market operations. These models have been employed to automate processes, forecast market trends, and empower the decision-making abilities of stakeholders. Some recent studies of stock market prediction using statistical, machine learning, and deep learning models are discussed in this section.

Zhang et al., 2019, introduced a new architecture for Generative Adversarial Networks (GANs) that utilizes a Multi-Layer Perceptron (MLP) as the discriminator and LSTM as the generator to forecast the closing prices of stocks. The model was evaluated by comparing it to other models, namely ANN, LSTMs, and SVR. The authors used data from the S&P 500 Index, Shanghai Composite Index, IBM, Microsoft Corporation, and Ping an Insurance Company of China (PAICC) over a 20-year period, which included opening, closing, lowest, and highest prices. The model was assessed using several evaluation metrics, including MAE, MAPE, RMSE, and Average Return (AR). The findings revealed that the proposed GAN model outperformed the other models [12].

Eapen et al., 2019, proposed a distinctive deep-learning model incorporating multiple pipelines of CNN and Bi-LSTM memory units for predicting the closing price for the S&P 500 index. The study used the historical records of the index spanning ten years, from January 2, 2008, to November 27, 2018. The dataset consisted of opening and closing stock prices and additional market parameters. The study's findings indicate that the proposed deep learning model outperforms the traditional SVM regressor regarding prediction accuracy from the temporal sequence data [13].

Alamro & Al-Rasheed, 2019, predicted the Saudi Stock Market Index by employing its historical data and considering the influence of individuals' sentiments on their financial decision-making processes. Media and news directly influence human emotions and opinions, which was incorporated using the Global Data on Events, Location, and Tone (GDELT) dataset. GDELT is a set of global news encompassing diverse media channels such as television, broadcasts, radio, newspapers, and websites. The authors analyzed the properties of the generated multivariate time series and subsequently implemented and evaluated various multivariate models: ARIMA, GRACH, and LSTM, to forecast the daily index of the Saudi stock market. The predicted RMSE was 0.61, while the MAE was 0.59 using LSTM, which showed discrepancies with the other methods [8].

Jarrah & Salim, 2019, predicted the trends in the stock prices of the Saudi stock market based on its historical prices by utilizing a combination of a recurrent neural network (RNN) and discrete wavelet transform (DWT). This study employed historical data on stocks from the Saudi stock market (Tadawul). About 190,000 series were collected in total from 2011/01/01 to 2016/03/31, with 1300 records for each company within the market. Each record kept a daily opening/ closing/ lowest/ highest/ volume. The acquired results were analyzed to compare them with the results obtained through traditional prediction algorithms such as the ARIMA. The proposed method achieved better MAE and RMSE values of 0.15996 and 0.19237 for a seven-day prediction. Similar to the research discussed in [10].

Pang et al., 2020, proposed two deep learning models: LSTM with an embedded layer and LSTM with an automatic encoder to predict the stock market prices. The data were collected from the livestock market for real-time and off-line analysis in the Shanghai stock exchange. In the two models, the authors operated the automatic encoder and the embedded layer, respectively, to vectorize the data to predict the stock using LSTM neural network. The outcomes of the experiment demonstrated the superiority of the deep LSTM with an embedded layer. For the Shanghai A-shares composite index, the accuracy of the two models was 57.2 and 56.9%, respectively. While for the individual stocks, the accuracy of the two models was 52.4 and 52.5%, respectively. The findings reflected the volatility of the Shanghai stock exchange prices [14].

Chen, 2020, investigated the analysis and forecasting of time series data. The authors selected four stocks from Yahoo Finance's historical database: Apple, Mastercard, Ford, and ExxonMobil. The dataset included the prices of opening, closing, daily high and low, adjusted closing, and trading volume from January 1st, 2002, to March 11th, 2020. Three machine-learning models were built for predicting future stock prices, including LSTM, CNN, and SVR. The MAPE was used to assess the performance of these models, and the results indicated that the combination of CNN and LSTM improved the accuracy of predictions compared to LSTM alone. Additionally, the SVR model predicted stock prices more accurately than the other two methods [15].

Kamalov, 2020, investigated the impact of previous changes on stock prices using machine learning algorithms. The study collected ten years of daily stock price data for four major United

States companies: Cisco, Nike, Coca-Cola, and Goldman Sacks. The author utilized three neural network architectures, CNN, MLP, and LSTM, to analyze the data and compare their performance against Relative strength index (RSI) and Random Forests models, which served as benchmarks. The models' performance was evaluated using the Area Under the Curve (AUC) metric. The results indicated that neural network models, particularly the LSTM model, can accurately predict significant changes in asset price. However, the study needed to provide adequate details on the data acquisition process and the features used in the models [16].

Nabipour et al., 2020, focused on predicting future trends in the stock exchange using data from the Tehran stock exchange. Specifically, the study examines data from four groups over ten years, from November 2009 to November 2019: petroleum, financials, basic metals, and non-metallic minerals. Different machine learning algorithms were employed to predict the future values of the stock market, including decision tree, bagging, random forests, adaptive boosting (AdaBoost), gradient boosting, eXtreme gradient boosting (XGBoost), as well as ANN, RNN, and LSTM. The performance of each prediction model was evaluated using four metrics: MAPE, MAE, RMSE, and MSE. The LSTM algorithm provides the most precise predictions with the highest model fitting ability. However, the authors should have elaborated on the rationale for selecting the four groups mentioned earlier for analysis [6].

Mehtab et al., 2021, proposed a hybrid modelling methodology to forecast stock prices utilizing various machine learning and deep learning-based models. The study employed the NIFTY 50 index values of India's National Stock Exchange from December 29, 2014, to July 31, 2020, with six features: the lowest, highest, opening price, closing price, the volume traded, and the range, computed as the difference between the highest and lowest prices. The authors applied SVM, random forests, decision trees, and LSTM with three different architectures to predict the opening price. The results indicated that the LSTM-based deep learning regression models significantly outperformed the machine-learning-based predictive models. The study's findings demonstrated that deep learning-based models could extract and learn time-series data features more effectively than their corresponding machine learning models [17].

Rashedi et al., 2021, employed a mathematical model using Artificial Intelligence to identify outliers in the daily return of the Saudi stock market. The dataset included closing price records from October 2011 to December 2019, obtained from the Saudi stock market, as well as repo rate, inflation rate, and oil prices. The repo rate and inflation rate were collected from the Saudi Authority for Statistics, while the oil prices were obtained from the Saudi Central Bank. The authors employed a combination of Radial Basis Function Neural Network (RBFNN) and particle swarm optimization algorithm to train their model. In this way, the authors classified the stock return into two categories: outliers and nonoutliers. The proposed model achieved a MSE value of 0.05 in categorizing the stock return data into outliers and nonoutliers. The proposed model used a powerful optimization approach to solve business procedure issues. However, their technique was used for outlier detection rather than regression [18].

Al-Nefaie & Aldhyani, 2022, investigated using the key study to predict a sector-adjusted closing price. Tadawul's historical records served as a basis for data collection for four sectors: communication, energy, financial, and industrial, from 2018 to 2020. These records included 12 features such as lowest and highest price, opening and closing price, change of stock market, number of trades, the volume traded, and information on companies listed on the exchange. The closing price is used to forecast the stock market's future value. Over a 60-day period, two machine learning algorithms were employed to predict future values, LSTM and MPL. The Pearson's correlation for these four sectors on the Saudi Stock Exchange were greater than 0.9950, indicating good outcomes [19].

A study by Aldhyani & Alzahrani, 2022, proposed a predictive model for stock market prices using financial time-series data as inputs. The model included LSTM and a hybrid of CNN with LSTM to predict the stocks of Apple and Tesla's closing prices. The Tesla dataset was collected from 4 August 2014 to 17 August 2017, while for Apple, it was collected from 3 January

2010 to 28 February 2020. The study assessed the model's performance using various evaluation metrics, including MSE, RMSE, NRMSE, and R. The results indicate that the CNN-LSTM model outperformed the single deep learning LSTM and existing systems in forecasting stock market prices [20].

The challenge of stock market prediction is because of the nature of the stock data: non-linear, dynamic, complicated [21], non-stationary characteristics of data [7] and the impact of economic conditions on market dynamics [1], [7]. Applying machine learning and deep learning approaches for prediction show the significant forecasting potential of these approaches compared with common approaches such as statistical analyses [7]. Thus, the aim of this study demonstrates how machine learning and deep learning methods outperform the traditional methods.

3. RESEARCH METHODOLOGY

This section reviews appropriate feature selection methods, appropriate models and algorithms that were applied to the dataset of the Saudi Stock Market as well as the evaluation methods.

3.1 Feature Selection

There are two techniques are used for numerical dependent and independent variables: Pearson's correlation and mutual information [22]. Pearson's correlation coefficient (R) is a measure of the linear relationship between two features; the R value lies between –1 and 1. A value of -1 indicates a negative correlation between the features, 1 indicates a positive correlation, and 0 indicates no correlation between the features [22].

$$R = \frac{N(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2]\,[N\Sigma y^2 - (\Sigma y)^2]}} \quad (1)$$

Mutual information is a statistical method used in feature selection, and it measures the amount of information that one random variable provides about another. Unlike traditional correlation measures, Mutual Information can capture linear and non-linear relationships between variables. To calculate the Mutual Information between two discrete random variables, Equation 2 is commonly used [23].

$$I(A,B) = \Sigma_{b\in B}\,\Sigma_{a\in A}\, p(a,b) log\left(\frac{p(a,b)}{p(a)p(b)}\right) \quad (2)$$

3.2 Prediction Models

The six models for predicting the stock market are as the following.

Autoregressive Integrated Moving Average (ARIMA) is a popular time series forecasting model used to make predictions based on historical data. It is a combination of three models: Autoregression (AR), Integrated (I), and Moving Average (MA) [23]. The AR component represents the use of past values of the dependent variable to forecast future values. The integrated (I) component represents the process of differencing the time series data to make it stationary, thus stabilizing the statistical properties of the series over time. The MA component involves using past forecast errors to predict future values [23]. The ARIMA model has three parameters: **p**, **d**, and **q**. The parameter p represents the number of lagged observations used in the model. The parameter d represents the degree of differencing applied to the raw observations to make the time series stationary. The parameter q describes the size of the moving average window, which is the order of the moving average used to calculate forecast errors. The model is built based on these parameters using a linear regression approach with the specified number and type of terms [24].

Support Vector Machine (SVM) models gains significant attention and popularity in academic research. SVM models are used for regression (SVR) and classification (SVC) problems [25]. In SVC, the goal is to find a hyperplane that can effectively classify data points. On the other hand, SVR is a kernel-based regression method that attempts to find the optimal regression hyperplane with the least structural risk in a high-dimensional feature space; This hyperplane is designed to accurately predict the values of a dependent variable for a given set of input features. SVR aims to minimize structural risk by finding the best hyperplane that maximizes the margin between the hyperplane and the closest data points, also known as support vectors [26]. The objective of the SVR is to find a function f(x) that minimizes the difference between the actual and predicted values within a specified error threshold ϵ. The problem is converted into a constrained optimization problem and can be mathematically represented using equation:

$$min \frac{1}{2} ||w||^2 + C . \sum_{i=1}^{n} (\varsigma_i + \varsigma_i^*) \\ s.t. \begin{cases} y - f(x_i) \le \varepsilon + \varsigma_i, \varsigma_i \ge 0 \\ y - f(x_i) \ge \varepsilon + \varsigma_i^*, \varsigma_i^* \ge 0 \end{cases} \quad (3)$$

In the above equation, the parameter C determines the trade-off between model complexity and the accuracy of sample fitting. The error threshold ϵ and slack variables are used to deal with outliers [25].

Random Forest is the most popular method for solving classification and regression problems. Random forests as an ensemble of multiple decision trees- grown on random subsets of the feature space; this randomness helps to reduce correlation among the trees and improves their independence, allowing the random forest to generalize better new data and overcome overfitting. The final prediction is obtained by averaging the predictions of all the individual trees in the forest [27], [28].

Long Short-Term Memory (LSTM) introduced by Hochreiter & Schmidhuber [29]; it is a special type of Recurrent Neural Network capable of learning long-term dependencies. LSTM is perfected over time to mitigate the long-term dependency issue that happens because of the vanishing gradient problem. Unlike traditional feedforward neural networks, LSTMs have feedback connections, allowing them to process entire sequences of data such as time series, text, and speech. They can retain information about previous data points in the sequence to help with the processing of new data, making them particularly effective at handling sequential data. The core concept of LSTMs is the cell state and its various gates. LSTMs use a cell state to carry information throughout the sequence, acting as the network's "memory." Gates are neural networks with sigmoid activations that control what information is added or removed from the cell state, allowing LSTMs to learn which information is relevant to keep or forget during training using the input gate and the forget gate. The cell state can retain information from earlier steps, reducing the impact of short-term memory. The output gate generates the next hidden state by combining the previous hidden state and current input using a sigmoid function [30], [31].

Bidirectional Long Short-Term Memory (Bi-LSTM) was created by Schuster and Paliwal and is a type of neural network that processes sequence data in both forward and backward directions [30]. Unlike the traditional LSTM, which processes the input sequence in a single direction, Bi-LSTM incorporates information from both past and future by using two separate hidden layers, one for the forward direction and one for the backward direction. Bi-LSTM is usually employed where sequence-to-sequence tasks are needed; this approach helps to capture the contextual information in both directions and can improve the accuracy of predictions in tasks such as text classification, speech recognition, and forecasting models [32].

Gated recurrent units (GRU) are the newer generation of RNNs and are similar to an LSTM but simplified in structure. GRU has fewer parameters and thus may train faster or need fewer data to generalize. It also has only two gates. (1) Update Gate: The update gate acts like the forget and input gate of an LSTM, and it decides what information to throw and what new information to keep. (2) Reset Gate: The reset gate is responsible for deciding how much of the previous hidden state to forget and how much new input to consider for the current hidden state. It is implemented as a sigmoid function, with a value of 1 indicating that the gate should let all information through and 0 indicating that it should block all information [33].

3.3 Evaluation Methods

The evaluation of the developed models was done using MSE, RMSE, MAE, and MAPE [34].

- Mean absolute error (MAE): MAE is a metric that calculates the average absolute differences between the predicted and actual values. The formula for MAE is as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \quad (4)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- Mean squared error (MSE): MSE is a metric used to determine the average squared distance between the actual and the predicted values. The formula for MSE is represented as:

$$MSE = \frac{\Sigma(y_i - \hat{y}_i)^2}{n} \quad (5)$$

- Root mean squared error (RMSE): RMSE is a commonly used metric for evaluating the accuracy of predictions by measuring the square root of the second sample moment of the differences between predicted and actual values. The formula for RMSE is represented as follows:

$$RMSE = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{n}} \quad (6)$$

- Mean Absolute Percentage Error (MAPE): MAPE is a metric used to measure the accuracy of a forecast by calculating the percentage difference between the predicted and the actual value. The formula for MAPE is represented as follows:

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{N} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (7)$$

4. DATA ANALYSIS AND IMPLEMENTATION

This section involves five main steps that were used. The first step was choosing and collecting the dataset. The second step exploring and visualizing the data. The third step was pre-processing the data. The fourth step was applying six prediction models. The last step evaluating the performance of these prediction models (see figure 1). The six involved sectors in this study were Banking, Basic Materials, Real Estate Management and Development, Insurance, Energy, and Telecommunication.

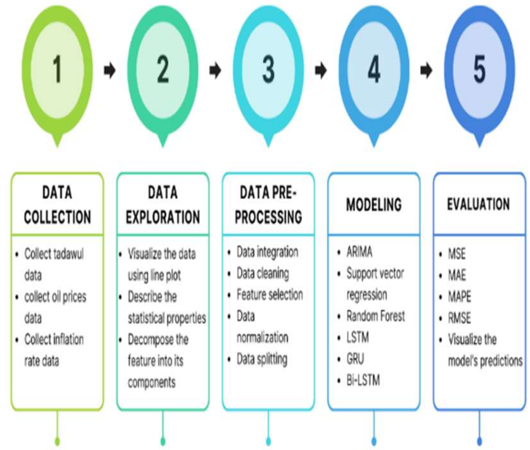

Figure 1: Steps For Predicting The Saudi Stock Market

4.1 Data Collection

The data collection step represents the dataset that was obtained from the Saudi Stock Exchange, known as Tadawul. These data comprised historical records for six sectors in the stock market from January 3, 2016, to December 25, 2022.

Table 1: Dataset Features

Feature	Description	Feature Type
Name of Sector	Name of the sector, which contains a specific type of companies	String
Date	Day of record	Date
Opening	Opening price for the stock on that day	Decimal
Highest	Highest price for the stock on that day	Decimal
Lowest	Lowest price for the stock on that day	Decimal
Closing	Closing price for the stock on that day	Decimal
Volume Traded	Volume of stocks traded on the day	Decimal
Value Traded	Value of stocks traded on the day	Decimal
Number of Trades	The number of trades in the day	Decimal
Average oil price	The average of oil price during a day	Decimal
Inflation rate	The value of the inflation rate during a month	Decimal

The sample size is 35598 observations. Each record contained the sector name, date, opening, closing, lowest and highest prices, traded volume and value, and the number of deals executed. Furthermore, the study incorporated two vital economic indicators, oil prices, and inflation rate, to examine their impact on the market. The data for these indicators were in the Trading Economics website, which provides access to over 20 million economic indicators across 196 countries [35]. The consisted of 11 distinctive features as shown in Table 1.

4.2 Data Exploration

The data exploration step aims to better understand the obtained data and its structure and identify any inherent patterns. For this purpose, the tasks were implemented:
(1) visualizing the data using line plots, (2) decomposing the data into its components, and (3) describing the statistical properties of the data. Figure 2 represent the closing price for all sectors listed in Tadawul using a line chart.

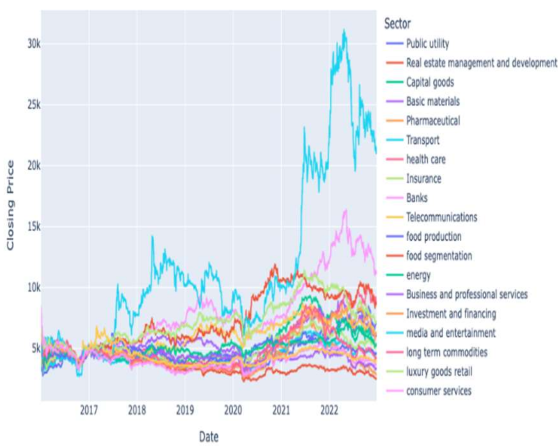

Figure 2: Closing Prices For All Sectors

The distribution of closing prices displays relatively similar patterns across most sectors, a slight upward trend observed after the 2020 period. It indicates that there are no significant differences in the behavior of prices among the sectors. However, it is worth mentioning that there is one sector (Media and Entertainment) that has shown a significant increase in prices during the last two years.

The process of selecting the sectors was done using the volume traded since the previous studies found that the volume traded significantly impacts the sectors' performance [36], [37]. Figure 3 displays the ordering of sectors based on the volume traded.

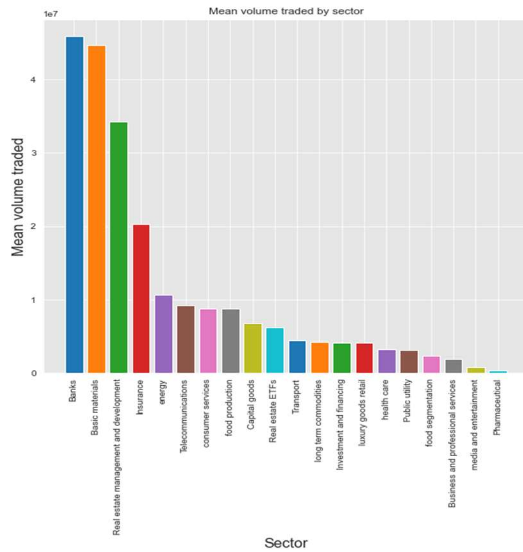

Figure 3: The Sectors Based on The Volume Traded.

The top six sectors were selected for further analysis. These sectors are Banking, Basic Materials, Real Estate Management and Development, Insurance, Energy, and Telecommunication as shown in Figure 4. This figure displays an upward trend over time, indicating a general growth in the stock market. At the same time, the magnitude of the closing price increased among the sectors, with some sectors exhibiting more rapid growth than others.

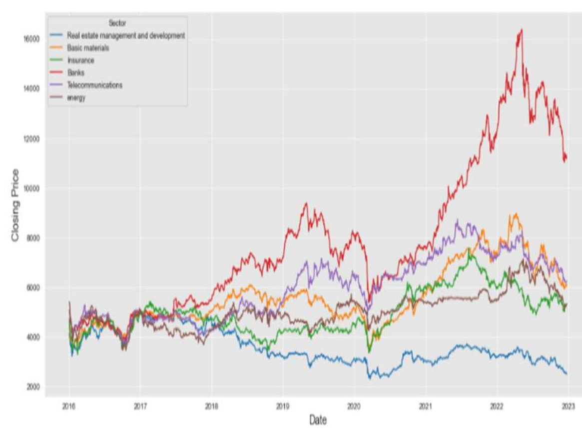

Figure 4: The Closing Price Of The Six Major Sectors

Figures 5 and 6 illustrate the oil prices and inflation rates. As shown in Figure 5, oil prices were highly volatile over time, with frequent fluctuations in prices. It also experienced a significant decline in the first half of 2020, coinciding with the outbreak of the COVID-19 pandemic and the subsequent

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

decline in global demand for oil. In Figure 6, the inflation rates had a relatively stable pattern time, with limited fluctuations observed in either direction.

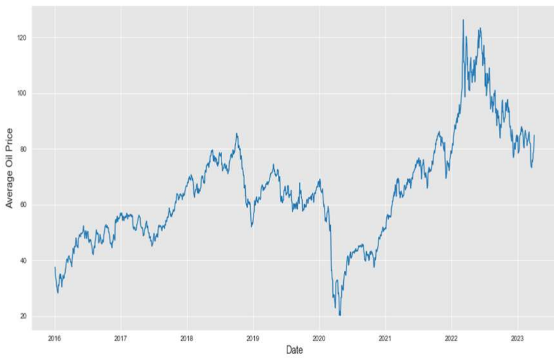

Figure 5: The Distribution Of The Average Oil Prices Overtime

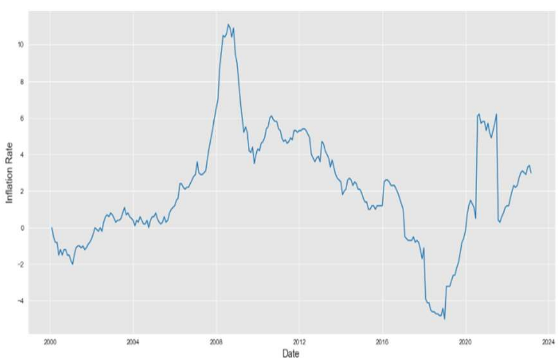

Figure 6: The Distribution Of Inflation Rates Overtime

The key descriptive statistics for selected sectors' features was applied. These statistics include opening and closing prices, highest and lowest prices, volume, and value traded, number of trades, as well as oil prices and inflation rates.

Tables from 2 to 7 represent descriptive statistics which include measures of central tendency, such as the mean and median. The mean represents the average value of a feature, while the median is the middle value when the dataset is sorted in ascending order. For example, the banking sector in table 2, the mean closing price is 7857.783, and the median (50th percentile) is 7118.920.

Standard deviation is another important descriptive statistic that measures the spread or dispersion of the data. For instance, the standard deviation for closing prices in the basic materials sector is 1,201.677 with a mean value of 5,606.603, suggesting a significant amount of dispersion in the closing prices for companies within that sector as shown in table 3.

Table 8: Summary Statistics For Oil Price And Inflation Rate

Feature	Average Oil Price	Inflation Rate
Min	75.675000	-5.000000
25%	190.200000	0.200000
50%	237.337500	1.700000
75%	282.862500	4.200000
Max	473.512500	11.100000
Mean	242.729692	2.029496
Standard Deviation	72.628256	3.130217

For oil prices and inflation rates, Table 8 also shows the descriptive statistics. Consider the average oil price percentiles as in the table. The minimum value indicates that the lowest value for the average oil price is 75.675. The 25th percentile value suggests that 25% of the data points in the average oil price had a value less than or equal to 190.2. The median or 50th percentile value signifies that 50% of the data points had a value less than or equal to 237.3375. The 75th percentile value indicates that 75% of the data points had a value less than or equal to 282.8625. The maximum value implies that at least 99% of the data points had a value lower than or equal to 473.5125.

4.3 Data Pre-processing

The pre-processing step was applied to prepare the data for further analysis as the following.

4.3.1 Missing values

In the context of time series data, missing values often occur in a pattern that may require a specific approach for handling. Therefore, forward imputation is a suitable technique for replacing the missing values with the most recently observed value for the same feature [38]. The dataset obtained from Tadawul has different dates from the oil prices dataset, as the Saudi Market trading days (Sunday to Thursday) are not aligned with global trading days (Monday to Friday). As a result, there were missing values that needed to be handled between the two datasets. To address this issue, a forward-filling approach was implemented. It involved filling the missing values with the most recent value for the same feature.

4.3.2 Feature Selection

Both Pearson correlation and mutual information techniques were used to reduce the number of input variables and identifying the most significant ones associated with the dependent variable, which was the closing price. The findings of these techniques are presented in Table 9.

Table 9: Pearson Correlation And Mutual Information Results

Feature	Pearson correlation	Mutual information
Opening	0.998355	3.201
highest	0.999709	4.464
lowest	0.999685	4.456
Volume traded	**-0.001656**	**0.129**
Value traded	**0.045684**	**0.147**
Number of traded	0.153810	0.155

The results demonstrate that volume and value traded variables exhibit relatively low correlation and mutual information values compared to the other variables. Therefore, these variables were excluded from the subsequent analysis.

4.3.3 Data integration

It involved combining multiple datasets into a single dataset to create a more comprehensive dataset for modelling. To incorporate the values of oil prices and inflation rates for each sector, a concatenation method was applied using Python to merge the three datasets (sector dataset, oil prices, inflation rate) based on the date column.

4.3.4 Data normalization

Data normalization is also known as data scaling transforms data into a standard range or distribution like Min-Max method to improve the performance of machine learning models [39]. The Min-Max normalization method is a linear transformation of the original data that scales the values of each feature to a range between 0 and 1. Min-Max normalization uses this formula:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (8)$$

Min-Max normalization was used, the normalized data is constrained to a range between 0 and 1, the influence of outliers on model performance was minimized, and the relative relationships between values within each feature were preserved.

4.3.5 Data Splitting

The dataset is typically split into training and test sets. The independent and dependent variables were defined as X and Y. where X included the opening, lowest, highest, oil price, and inflation rate values, and Y contained the closing price. Then, X and Y were split into training and testing sets with a ratio of 90:10, where 90% of the data is used for training, and the remaining 10% is used for testing. As the data is time-series, the first 90% of the data is used for training, while the last 10% is used for testing.

4.4 Implementation Models

In the implementation step, six Models were implemented to predict the closing price of six major sectors in the Saudi stock market. These models were statistical model ARIMA, two machine learning models (SRV and Random Forest) and three deep learning models (LSTM, Bi-LSTM, and GRU).

ARIMA: Initially, to determine the optimal values for the ARIMA model parameters (p, d, q), the 'auto_arima' function was utilized. After finding the best parameters, which were determined to be ARIMA (0,1,0) (0,0,0) [0], the model was implemented using these parameters.

SVR: The GridSearchCV function was used. The parameters included C, gamma, epsilon, and kernel, and a set of defined values for each parameter. The best parameters were determined to be C = 1, epsilon = 0.01, gamma = 0.1, and kernel = linear.

Random Forests: The GridSearchCV function was used to determine the optimal hyperparameters for the Random Forests model. The parameter space explored during the search included three hyperparameters: n_estimators, max_depth, and

Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

min_samples_split. The parameters were max_depth = 15, min_samples_split = 5, and n_estimators = 300.

Deep learning-based models: Three deep learning models: LSTM, Bi-LSTM, and GRU, were built. The architecture of these models included several parameters that are used to configure the model and influence its performance. These parameters included the number of neurons, layers, activation function, drop rate, optimizer, and epochs. The activation function is used to introduce non-linearity to the model and transform the output of the neurons to the desired range. Tangent hyperbolic (tanh) has been used, a mathematical function that maps the input values to a range between -1 and 1. Tanh calculated as follows:

$$f(x) = \frac{(e^{x} - e^{-x})}{(e^{x} + e^{-x})} \tag{9}$$

Drop rate was a regularization technique used to prevent overfitting by randomly dropping out input units during training. For LSTM, drop rate is 0.0 indicated that no units will drop during the training, while for GRU and Bi-LSTM 0.2 of the units dropped when moving from layer to another. RMSprop optimizer used for LSTM while Adam optimizer used for GRU and Bi-LSTM, both known for handling large datasets efficiently and adapting the learning rate for each weight individually. Finally, epochs for the three models were 1000.

4.5 Evaluation

In evaluation step, four evaluation metrics MSE, RMSE, MAE, and MAPE were used to evaluate the implemented models for the six sectors. These sectors were Banking, Basic Materials, Insurance, Real estate management & development, Energy, and Telecommunication.

5. RESULTS AND DISCUSSION

This section discusses the results of the implemented the six prediction models for the six sectors.

5.1 *Energy Sector*

Table 10 and figure 7 represent the results for the developed models in the energy sector. The findings reveal that the SVR model outperforms the other models, with the GRU network model following closely.

Table 10: Evaluation Results For Energy Sector

	MSE	RMSE	MAE	MAPE
ARIMA	8660.07595	93.05953	76.74758	1.29322
SVR	**1551.49404**	**39.38901**	**31.79364**	**0.50682**
RF	2776.89439	52.69625	42.30313	0.67345
LSTM	4467.74782	66.84121	56.08252	0.88558
Bi-LSTM	2016.60876	44.90667	34.33633	0.54134
GRU	1654.51049	40.67567	33.80273	0.54171

Figure 7: Models Performance For Energy Sector

5.2 Banking Sector

The evaluation results for the developed models in the banking sector are presented in Table 11 and figure 8. The outcomes show that the machine learning models (Random Forests followed by SVR) have performed better result than the other models.

Table 11: Evaluation Results For Banking Sector

	MSE	RMSE	MAE	MAPE
ARIMA	193914.424230	440.35716	396.43657	3.07248
SVR	22710.70779	150.70072	120.79499	0.88660
RF	**22655.08206**	**150.516052**	**115.98496**	**0.84816**
LSTM	71415.41939	267.23663	237.02709	1.79127
Bi-LSTM	41883.35526	204.65423	165.01172	1.24375
GRU	55398.77634	235.36945	203.38363	1.48454

Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 8: Models Performance For Banking Sector

5.3 Basic Materials Sector

Table 12 and figure 9 demonstrate the performance results for the Basic Material sector. The results indicate that the GRU model outperforms the other models. However, it is noteworthy that all machine and deep learning models have obtained similar results that differ from the ARIMA model, which is considered the worst.

Table 12: Evaluation Results For Basic Materials Sector

	MSE	RMSE	MAE	MAPE
ARIMA	60592.99252	246.15644	226.30969	3.34032
SVR	3867.55688	62.18968	48.53877	0.68826
RF	3312.95347	57.55826	44.24551	0.62351
LSTM	6856.12838	82.80174	65.12595	0.89860
Bi-LSTM	4339.19249	65.87255	51.86241	0.73424
GRU	**3242.82816**	**56.94583**	**44.40341**	**0.62261**

Figure 9: Models Performance For Basic Materials Sector

5.4 Telecommunication Sector

Regarding the telecommunication sector, the evaluation outcomes demonstrate that the GRU has attained superior performance compared to the other models, followed by Bi-LSTM as shown in table 13 and figure 10.

Table 13: Evaluation Results For Telecommunication Sector

	MSE	RMSE	MAE	MAPE
ARIMA	44088.65831	209.97299	196.39641	2.87009
SVR	5164.94198	71.86753	60.06227	0.86167
RF	5581.69602	74.71075	63.15062	0.90483
LSTM	6442.43614	80.26479	66.16829	0.94025
Bi-LSTM	4792.00430	69.22430	58.92342	0.84088
GRU	**4773.65283**	**69.09162**	**58.72284**	**0.83685**

Figure 10: Models Performance For Telecommunication Sector

5.5 Real Estate Management and Development Sector

The table 14 and figure 11 indicate that the Random Forests model has achieved superior performance in the Real estate management and development sector compared to other models. Meanwhile, the GRU and SVR models have demonstrated similar results. In contrast, the ARIMA model has performed poorly compared to the other models.

Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 14: Evaluation Results For Real Estate Management And Development Sector

	MSE	RMSE	MAE	MAPE
ARIMA	1462.91297	38.24805	30.37751	1.00615
SVR	538.44093	23.20433	18.05951	0.59430
RF	**513.83322**	**22.66789**	**16.68420**	**0.54752**
LSTM	640.95388	25.31707	19.62951	0.65178
Bi-LSTM	648.38696	25.46344	20.46562	0.66547
GRU	535.86797	23.14882	18.55808	0.61088

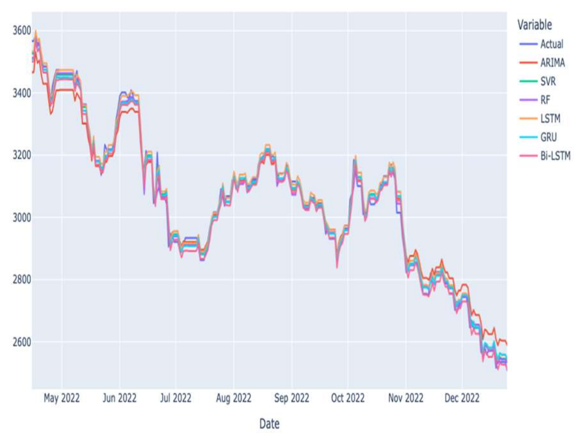

Figure 11: Models Performance For Real Estate Management And Development Sector

5.6 Insurance Sector

Table 15 and figure 12 demonstrate that the Bi-LSTM model achieved the best performance among the other models and the second model is SVR, while the ARIMA model exhibited the worst performance.

Table 15: Evaluation Results For Insurance Sector

	MSE	RMSE	MAE	MAPE
ARIMA	4210.05764	64.88496	53.41559	1.00349
SVR	1851.46128	43.02860	33.96688	0.62739
RF	1945.13522	44.10369	34.35516	0.63517
LSTM	3245.00659	56.96496	46.11299	0.85942
Bi-LSTM	**1568.15884**	**39.59999**	**31.53461**	**0.58283**
GRU	2028.77660	45.04194	35.91683	0.66647

Figure 12: Models Performance For Insurance Sector

Overall, the evaluation outcomes reveal that the GRU and Random Forests models have demonstrated superior performance among the other models across most sectors. Moreover, the SVR and Bi-LSTM models have exhibited promising results, while the LSTM model has mediocre performance. On the other hand, the ARIMA model has persistently underperformed across all sectors, indicating its unsuitability for forecasting stock prices.

5.7 Comparing with Previous Studies

It is important to compare this study's findings with previous studies. Therefore, a comparison of the study results with two previous studies is presented in Table 16.

Table 16: Comparing The Proposed Models With The Previous Studies.

Dataset	Model	RMSE	MAPE	MSE
India's National Stock Exchange	LSTM	344.57		
	Random Forests	0.42		
Tehran stock exchange	LSTM		0.54	225,333.35
	Random Forests		1.36	5,192,173.88
Tadawul + oil prices & inflation rate	LSTM	25.31707	0.65178	640.95388
	Random Forests	22.66789	0.54752	513.83322

In this table, Two previous studies, Tehran stock exchange [6] and India's National Stock Exchange [17], were selected for comparison with this research study due to their similarity in terms of data and Study by [17] assessed the effectiveness of

Random Forests and LSTM models in predicting the NIFTY 50 index, which consists of 50 major companies in India's National Stock Exchange. The performance of these models was evaluated using the RMSE metric. The results of proposed LSTM model achieved an RMSE of 25.31707, outperforming the LSTM model examined in the study of India's National Stock Exchange, which obtained an RMSE of 344.57. Similarly, in the study by [6], the Random Forests and LSTM models were applied in four sectors in the Tehran stock exchange, and the MAPE and MSE metrics were used for evaluation. The results clearly indicated that the proposed models (LSTM and Random Forests) outperformed the models examined in the study of Tehran stock exchange [6]. Specifically, the results of the proposed models achieved 0.65178 and 0.54752 of MAPE and 640.95388 and 513.83322 of MSE, while in the study of Tehran stock exchange, the MAPE values were 0.54 and 1.36 and the MSE were 225,333.35 and 5,192,173.88, respectively.

6. CONCLUSION

Investors and decision-makers in the financial market continually strive to predict stock market changes effectively, enabling them to make informed investment decisions. Therefore, each stock market must possess robust tools and techniques that provide stakeholders valuable information and insights into market fluctuations. The, the contributions of this research are: (1) the fusion of trading information and economic factors to improve the prediction (2) using feature extraction to extract useful trading information and (3) applying deep learning to predict accurately stock market prices. The closing prices were predicted for six major sectors in the Saudi stock market (Banking, Basic Materials, Real Estate Management and Development, Insurance, Energy, and Telecommunication), using prediction models. These models were ARIMA, SVR, Random Forests, LSTM, Bi-LSTM, and GRU. The study incorporated two important economic indicators (oil prices and inflation rate) to analyze their impact on the stock market.

The findings revealed that GRU and Random Forests demonstrated superior performance compared to the other models in accurately predicting stock prices across the analyzed sectors. GRU achieved a 3242.82816 MSE, 56.94583 RMSE, 44.40341 MAE, and 0.62261 MAPE, while Random Forests achieved a 513.83322 MSE, 22.66789 RMSE, 16.68420 MAE, and 0.54752 MAPE. Moreover, the SVR and Bi-LSTM models have exhibited promising results, SVR achieved MSE of 1551.49404, RMSE of 39.38901, MAE of 31.79364, and MAPE of 0.50682, and Bi-LSTM achieved MSE 1568.15884, RMSE of 39.59999, MAE of 31.53461, and MAPE of 0.58283. On the other hand, the LSTM model exhibited mediocre performance, while the ARIMA model consistently performed poorly. This study emphasizes the potential of advanced machine learning techniques in predicting stock prices and underscores the importance of employing robust methodologies for accurate predictions.

Future work could explore hybrid techniques that leverage the strengths of different models to provide more comprehensive and accurate predictions. Stock prices are influenced by various factors such as economic, social, and political and incorporating them in the forecasting could enhance the predictive power of stock market.

REFERENCES

[1] Nti I., Adekoya A., Weyori B. 2019 A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review. 53(4). PP. 3007–3057.

[2] G.C. Surya Bahadur 2006 Stock market and economic development: A causality test. Journal of Nepalese Business Studies. 3(1). PP. 36–44. Available at SSRN: https://ssrn.com/abstract=1566002.

[3] Hu Z., Liu W., Bian J., Liu X., and Liu T. 2018. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. ACM, Los Angeles, California, USA, PP. 261–269.

[4] Zheng J., Wang Y., Li S. and Chen H. 2021 The stock index prediction based on SVR model with bat optimization algorithm. Algorithms. 2021;14(10):299.

[5] Akhter S., Misir M. 2005 Capital markets efficiency: evidence from the emerging capital

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

market with particular reference to Dhaka stock exchange. South Asian Journal of Management. 12(3). PP. 35-51.

[6] Nabipour M., Nayyeri P., Jabani H, Mosavi A., Salwana E., Shahab S. 2020 Deep learning for stock market prediction. Entropy. 22(8): 840. PP. 2-23.

[7] Kumar D., Sarangi P., Verma R. 2021 A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings 49(1).

[8] Alamro R, McCarren A, Al-Rasheed A. 2019 Predicting Saudi stock market index by incorporating gelt using multivariate time series modelling. In International Conference on Computing Dec 10. Springer, Cham. PP. 317-328.

[9] Saudi Stock Exchange, 2023. [Online]. Available: https://www.saudiexchange.sa/wps/portal/saudiexchange/newsandreports/reports-publications/historical-reports?locale=en. [Accessed: 16-Jan-2023].

[10] Jarrah M, Salim N. 2019 A recurrent neural network and a discrete wavelet transform to predict the Saudi stock price trends. International Journal of Advanced Computer Science and Applications. 10(4).

[11] Chong, Eunsuk & Han, Chulwoo & Park, Frank. (2017). Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies. Expert Systems with Applications. Vol. 83(4). PP. 1-36.

[12] Zhang K., Zhong G., Dong J., Wang S., Wang Y. "Stock market prediction based on generative Adversarial Network", Procedia Computer Science. 147, 2019, PP. 400–406

[13] Eapen J., Bein D., Verma A. "Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction", 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC).

[14] Pang X, Zhou Y, Wang P, Lin W, Chang V. "An innovative neural network approach for stock market prediction", The Journal of Supercomputing. 76(3), 2020 PP. 2098-2118.

[15] Chen L-P. "Using machine learning algorithms on prediction of stock price", Journal of Modeling and Optimization, Vol.12(2), 2020, PP. 84–99.

[16] Kamalov F. "Forecasting significant stock price changes using neural networks", Neural Computing and Applications, 32(23), 2020, PP. 17655–17667.

[17] Mehtab S., Sen J., Dutta A. "Stock price prediction using machine learning and LSTM-based deep learning models". In: Thampi, S. M., Piramuthu, S., Li, K. C., Beretti, S., Wozniak, M., and Singh, D. (eds.), Machine Learning and Metaheuristics Algorithms and Applications. SoMMA 2020. Communications in Computer and Information Science, Vol 1366, 2021, pp. 88-106, Springer, Singapore. DOI: 10.1007/978-981-16-0419-5_8.

[18] Rashedi K., Ismail M., Hamadneh N., AL Wadi S., Jaber J., Tahir M. "Application of radial basis function neural network coupling particle swarm optimization algorithm to classification of Saudi Arabia stock returns", Journal of Mathematics. Vol. 2021, PP.1-8. https://doi.org/10.1155/2021/5593705.

[19] Al-Nefaie AH., Aldhyani TH. "Predicting Close Price in Emerging Saudi Stock Exchange: Time Series Models", Electronics, Vol. 11(21), 2022, PP. 1-25, DOI:10.3390/electronics11213443.

[20] Aldhyani T.H. and Alzahrani A. "Framework for predicting and modelling stock market prices based on deep learning algorithms", Electronics. Vol. 1(19), 2022, 3149. DOI:10.3390/electronics11193149.

[21] Tan, T. Z., Quek, C., & Ng, G. S. "Brain inspired genetic complimentary learning for stock market prediction". In IEEE congress on evolutionary computation, 2–5th September, Vol. 3, 2005, pp. 2653–2660.

[22] Venkatesh B. and Anuradha J. "A review of feature selection and its methods", Cybernetics and Information Technologies, Vol. 19(1). 2019, PP. 3–26.

[23] Wirawan I.M. and Widiyaningtyas T. and Hasan MM. "Short term prediction on Bitcoin Price using Arima method", International Seminar on Application for Technology of Information and Communication (iSemantic), 2019.

[24] Makala D. and Li Z. "Prediction of gold price with Arima and SVM. Journal of Physics", Conference Series. 767(1), 2021, 012022.

[25] Zhang J, Teng Y-F, Chen W. "Support vector regression with modified Firefly algorithm for stock price forecasting", Applied Intelligence, 49(5), 2018, PP.1658–1674.

[26] Basak S., Kar S., Saha S. Khaidem L. and Dey SR. “Predicting the direction of stock market prices using tree-based classifiers”, The North American Journal of Economics and Finance. 2019; 47, PP. 552–67.

[27] Zhang J., Cui S., Xu Y., Li Q. and Li T. “A novel data-driven stock price trend prediction system”, Expert Systems with Applications, 97, 2018, PP. 60–69.

[28] Hochreiter S. and Schmidhuber J. “Long short-term memory”, Neural Computation, Vol. 9(8), 1997, PP.1735–1780.

[29] Sunny A. Maswood M. and Alharbi A. “Deep learning-based stock price prediction using LSTM and bi-directional LSTM model”, 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020, DOI:10.1109/NILES50944.2020.9257950.

[30] Ghosh A., Bose S., Maji G., Debnath N., and Sen S. “Stock price prediction using LSTM on Indian share market”, EPiC Series in Computing. Vol. 63, 2019, PP. 101–110.

[31] Kwak G. Park C. Ahn H. Na S. Lee K. and Park N. “Potential of Bidirectional Long Short-Term Memory Networks for Crop Classification with Multitemporal Remote Sensing Images”, Korean Journal of Remote Sensing [Internet], Vol. 36(4). 2020, PP. 515–525, https://doi.org/10.7780/KJRS.2020.36.4.2.

[32] Hamayel MJ. and Owda AY. “A novel cryptocurrency Price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms”, AI. Vol. 2(4). 2021, PP. 477–496.

[33] Saud AS, Shakya S. “Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE”, Procedia Computer Science. Vol. 167(2), 2020, PP. 788-798.

[34] KS R. and MK S. “Stock market prediction using Deep Learning Techniques”, International Conference on Communication, Control and Information Sciences (ICCISc), Idukki, India. 2021, PP. 1-6. DOI: 10.1109/ICCISc52257.2021.9484960.

[35] Trading Economics, 2023. [Online]. Available: https://tradingeconomics.com/. [Accessed: 16-Apr-2023].

[36] Hussain S, Jamil H, Javed M, Ahmed W. “Analysis of relationship between stock return, trade volume and volatility: Evidences from the banking sector of Pakistani market”. European Journal of Business and Management, Vol. 6(20), 2014, PP. 57-61.

[37] Krishnamurthy A, Suresha B. “A study on the performance of Agro based stocks-An Evidence from select listed companies”. Asian Journal of Management. 2019, Vol. 10(4), PP. 389-393.

[38] Suo Q. Yao L. Xun G. Sun J. and Zhang A. “Recurrent imputation for multivariate time series with missing values”, 2019 IEEE International Conference on Healthcare Informatics (ICHI).

[39] Singh D. and Singh B. “Investigating the impact of data normalization on classification performance”, Applied Soft Computing. 97, 2020, 105524.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Statistics For All The Features Of Banking Sector

Feature	Min	25%	50%	75%	Max	Mean	Standard Deviation
Opening	0	5335.330	7120.670	9181.355	16,499.05	7854.234	3015.207
Highest	3805.320	5358.500	7157.290	9265.380	1670.059	7908.490	3040.997
Lowest	3805.320	5307.210	7069.500	9122.740	16,317.18	7804.147	2973.695
Closing	3805.320	5337.545	7118.920	9194.375	16,389.68	7857.783	3005.437
Volume traded	5243667	27882920	39191900	55777000	436133400	45966070	28,978,240
Value traded	219907800	668215900	866081700	1161995000	13804470000	1025091000	723372100
Number of traded	2704	9361	15113.00	25406.00	111803.0	19693.48	14311.52

Table 3: Statistics For All The Features Of Basic Materials Sector

Feature	Min	25%	50%	75%	Max	Mean	Standard Deviation
Opening	0	4,758.67	5,296.9	6,168.655	8,972.37	5,605.306	1211.485
Highest	3,489.8	4,783.55	5,333.41	6,211.685	8,997.96	5,637.13	1,215.618.
Lowest	3,463.19	4,729.22	5,255.050	6,108.395	8,955.8	5,574.274	1189.433
Closing	3,466.47	4,758.67	5,296.90	6,172.575	8,980.310	5,606.603	1,201.677
Volume traded	8072850	26808810	36959940	52530170	2148329000	44651930	55837710
Value traded	286877900	762871200	1078661000	1558046000	260000000000	1406853000	6237632000
Number of traded	6555	20018.50	30856.00	66095.50	226111.00	43995.11	30345.22

Table 04: Statistics For All The Features Of Insurance Sector

Feature	Min	25%	50%	75%	Max	Mean	Standard Deviation
Opening	0	4276.61	4744.34	5588.025	7632.38	4967.107	890.2066
Highest	3288.43	4293.27	4768.19	5636.45	7673.5	5001.57	889.9361
Lowest	3288.43	4249.365	4723.46	5533.44	7405.7	4937.302	872.9527
Closing	3288.43	4276.985	4744.34	5579.52	7574.24	4969.021	881.3199
Volume traded	2733682	9119254	14235660	26519060	138466900	20284990	16387230
Value traded	48703090	194160100	302354200	555741100	3720026000	447337862	400506800
Number of traded	4466	13481	18516	27208.5	132579	22815.42	14656.76

Table 05: Statistics For All The Features Of Real Estate Management And Development Sector

Feature	Min	25%	50%	75%	Max	Mean	Standard Deviation
Opening	0	3114.945	3362.98	4180.68	5058.93	3590.875	680.8021
Highest	2371.83	3133.93	3382.99	4203.565	5091.24	3613.621	675.1524
Lowest	2285.43	3089.435	3332.8	4153.87	5058.93	3570.22	675.4123
Closing	2294.42	3112.84	3360.85	4179.025	5058.93	3590.997	675.5576
Volume traded	2999627	14631780	26127360	44928250	270694100	34259610	28345820
Value traded	41657340	180866700	312982000	523200800	3428961000	397604500	305144600
Number of traded	2059	6239	11912	19528.5	97700	13960.78	9680.498

Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 0: Statistics For All The Features Of Energy Sector

Feature	Min	25%	50%	75%	Max	Mean	Standard Deviation
Opening	0	4,539.2	4,963.34	5,485.77	7,137.44	5,026.492	710.5253
Highest	3,459.53	4,569.495	4,988.03	5,509.855	7,278.96	5,058.637	707.2805
Lowest	3,459.53	4,502.625	4,932.97	5,459.77	7,087.52	4,998.175	693.2061
Closing	3,459.53	4,540.17	4,963.52	5,487.26	7,136.56	5,029.762	699.5980
Volume traded	810223	3844655	7699419	13303930	464,523,800	10,640,440	15,054,890
Value traded	19642370	79769310	174664100	389590990	15978620,000	305105700	519892500
Number of traded	720	2499	4879	16234.5	149454	10446.76	10,987.82

Table 7: Statistics For All The Features Of Telecommunication Sector

Feature	Min	25%	50%	75%	Max	Mean	Standard Deviation
Opening	0	4,843.18	6,023.76	6,965.34	8,756.10	5,966.59	1,252.37
Highest	3,717.55	4,876.59	6,065.35	7,001.64	8,796.24	6,008.79	1,257.89
Lowest	3,717.55	4,816.38	5,973.76	6,916.30	8,608.83	5,926.71	1,227.64
Closing	3,717.55	4,843.88	6,028.28	6,960.56	8,739.95	5,968.16	1,241.81
Volume traded	0	4769974	7,330,479	11,426,044	307,684,800	9,161,668	9,726,056
Value traded	0	85938070	155827600	260589500	12319550000	209048400	346228700
Number of traded	0	2697.50	4901	8697	41920	6436.09	5111.28