Enhancing Stock Price Prediction Using Mutual Information, PCA, and LSTM: A Deep Learning Approach

Zinah Kareem Mansoor[1,*], Ali Yakoob Al-Sultan[1]

[1]Department of Computer, College of Science for Women, University of Babylon, Babylon, Iraq

Emails: scw734.zynh.kareem@student.uobablyon.edu.iq; ali.alsultan@uobabylon.edu.iq

Abstract

The stock price exhibits quick and extremely nonlinear fluctuations in the financial market. A prominent worry among scholars and investors is the correct prediction of short-term stock prices and the corresponding upward and downward trends. Financial organizations have successfully incorporated machine learning and deep learning techniques to anticipate time series data accurately. Nevertheless, the precision of these models' predictions still needs improvement. Most current studies employ single prediction algorithms that cannot overcome intrinsic limitations. This paper proposes a methodology that utilizes the MUTUAL, principal component analysis (PCA), and Long Short-Term Memory (LSTM) model to accurately simulate and predict the variations in stock prices. The technology is utilized for the three global stock market datasets: TSLA, S&P500, and NASDAQ. The highest level of improvement achieved is a correlation of 99%. Furthermore, there is a reduction in error for the metrics MSE, MAPE, and RMSE, with improvements of 0.0001, 0.009, and 0.01 correspondingly.

Keywords: PCA; LSTM; Deep learning; Stock Market Prediction, Deep Learning, Neural Networks, Sentiment Analysis, Financial Markets, Trading Strategies

1. Introduction

The stock market can be described by volatility, unpredictability and the fact that it is a non-linear system. It must be understood that stock price forecasting is not an easy task as it involves many factors like the political conditions, world economy, and the company's financial statements and results. For this reason, the firm needs to adopt various techniques that enhance the revenues as well as the reduction of expenses in the course of realizing the company's goals and objectives. Therefore, to maximize profits and minimize such loses; there will be a greater benefit to employ strategy that tries to forecast future stock prices for the firm based on previous data trends over the few years thus helping in addressing the issues of volatility of the market. [1] [2]. Prevailing literature on Stock price prediction has highlighted two key approaches that have been traditionally used to predict a corporation's stock price. The technical analysis method involves forecast of market price of a particular stock by analyzing the historical records of closing and opening prices of the stock, trading volume of the particular stock, tendency in the nearby days in terms of closing prices. The second type of analysis is categorical and, like the first type, depends on external factors such as the company itself, the market, including economic and political factors, financial information from news, articles, posts on social networks and blogs, and other economic materials [3].

Nowadays, sophisticated intelligent methods that rely on either technical or fundamental analysis are employed to forecast stock values. Specifically, when it comes to analyzing the stock market, the amount of data is substantial and exhibits non-linear patterns. An efficient model is required to handle this diverse range of data, capable of identifying the concealed patterns and intricate relationships within this extensive dataset. Machine learning algorithms in this field have demonstrated a significant increase in efficiency, ranging from 60 to 86 percent, when compared to traditional methods [4].

Most of the previous work in this area like Decision Tree (DT) [5], Support Vector Machine (SVM) [6], Random Forest (RF) [7] . Furthermore, studies indicate that deep learning models, such as (LSTM)[8], exhibit superior performance compared to machine learning models, such as support vector regression (SVR) [9]. Different types of machine learning algorithms present with diverse strengths in analyzing the same historical data. Performance

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

depends on the nature of the data itself and the time duration of the term that the historical data is at hand. An increasingly developing area of financial and statistical studies which use different machine learning algorithms to reconstruct financial time series data for performing stock market price forecasts and performance evaluation [6]. Machine learning and deep learning advancements have opened up new possibilities for constructing stock price prediction models using time-series data that exhibit high cardinality. These models are used to predict future fluctuations in stock prices[10][11].

Three approaches are employed in this study including the Mutual Information, Principal Component Analysis, and the Long Short Term Memory. New variables generated from the financial data set are employed in the model such as the Open, High, Low, and Close values of a particular firm. These extra indicators will therefore be useful in improving the accuracy of the models, in predicting the closing price of a specific firm for the next day. The performances of the developed models are measured and tested using the Mean Squared Errors (MSEs), Root Mean Squared Errors (RMSEs), and Mean Absolute Percentage Errors (MAPEs). The remaining papers are as follows: Section 2 examines the existing research in the field. Section 3 outlines the methodology and procedures that were used. Section 4 presents the findings, while section 5 provides the conclusion the study.

2. Related Work

Several academics have researched the stock market to develop ideas for its functioning. The stock market known for its unpredictable nature and complex workings, for this problem, deep learning and machine learning techniques have been used. Most of them are described below.

(Iyyappan et al.2022)[12] set an objective to develop an AI platform for trend forecasting of stock market using advanced machine learning models. A further addition to the study was the LSTM, which was used with the research. In many cases, the average RMSE was below 50, which is an indication that the method they adopted was effective. (Kelany, Aly, and Ismail 2020)[13] studied the future prices of stock market sectors using a variety of machine learning techniques. The LSTM model achieved a low MAPE value but a somewhat long runtime, indicating a trade-off between accuracy and computing economy. (Mazed 2019)[14] studied the performance of three stock price prediction models: ARIMA, PROPHET, and KERAS with LSTM. These models were built and compared using historical stock price data from the National Stock Exchange (NSE). The results showed that all of the models performed better when applied to larger datasets. Notably, the LSTM model was the most accurate in forecasting stock price. (Patel and Saket 2018)[15] used the LSTM model to anticipate stock prices based on historical data, taking into account parameters such as price terms (e.g., starting and closing prices) and trading volume. The researchers performed a thorough examination of the data, visualizing the anticipated price values over time and determining the best parameters for the model. The model's final test result was given as 0.69205044 Mean Squared Error (MSE) and 0.83189569 Root Mean Squared Error. (Wen, Lin, and Nie 2020)[16] are used. PCA-LSTM predictive model. Using Pingan Bank's stock price as an example, PCA extracts features to provide crucial technical indices that influence stock price, while LSTM predicts stock price. The experimental findings revealed that the PCA-LSTM model had an RMSE of 0.221 and a MAPE of 1.667%. (Bathla 2020)[17] Applied Long Short-Term Memory (LSTM) to predict fluctuations in stock values. In the experiment, the two methods LSTM and SVR were compared using different stocks index data including S&P 500, NYSE, NSE, BSE, NASDAQ and DJIA. The result of the test demonstrates that LSTM, in fact, is able to reach a significant level of accuracy. (Gurav and Sidnal 2018)[18] introduced an adaptive stock forecasting model by proposing a unique modified back propagation neural network. Experiments are carried out for diverse markets. The MBNN achieves excellent success across all markets. The MBNN model achieved an average improvement in RMSE performance of 26.93%. (Song, Chung Baek, and Kim 2021)[19] proposed one of the most effective approaches of denoising which employed Fast Fourier Transform (FFT) inclusive of padding to pare noise. The effectiveness of the proposed denoising technique was confirmed through experiments that applied it to the basic indicators of several countries and calculated the stock index of the subsequent day. The findings also show that the integration of deep learning models with the proposed denoising strategy outperforms the basic models not only in terms of prediction but also eradicates the problem of time delay. (Mohan et al. 2019)[20] is used time series models, artificial neural networks, and a combined model of artificial neural networks and financial text data. The result shows a strong positive relation between stock prices and the selected financial news stories. . The RNN time series-forecasting model has demonstrated superior performance in predicting stock price direction, revealing a significant association between textual information and stock price movement. The models exhibited weak results in instances where stock prices were low or exhibited excessive volatility.

(Arefin 2021)[21] Is utilized machine learning technique to establish a price predicting system for second hand Tesla vehicle. To achieve this goal, used different machine learning techniques including decision trees, support vector machine (SVM), the Boosted Decision Tree algorithm achieved the lowest RMSE.

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

3. Methodology

The prediction analysis is a kind of statistical, modeling, data mining and some kind of machine learning to analysis the historical data and from the analysis, data miner can get a power to make further prediction about the information. SM analysis requires analyzing the historical stock price data of a company to get a rough estimate about the future of the stock price. Predict in this study involve the following Models: MUTUAL-PCA- LSTM Models employing Financial Deep Learning Neuron Network for stock price prediction in the future.

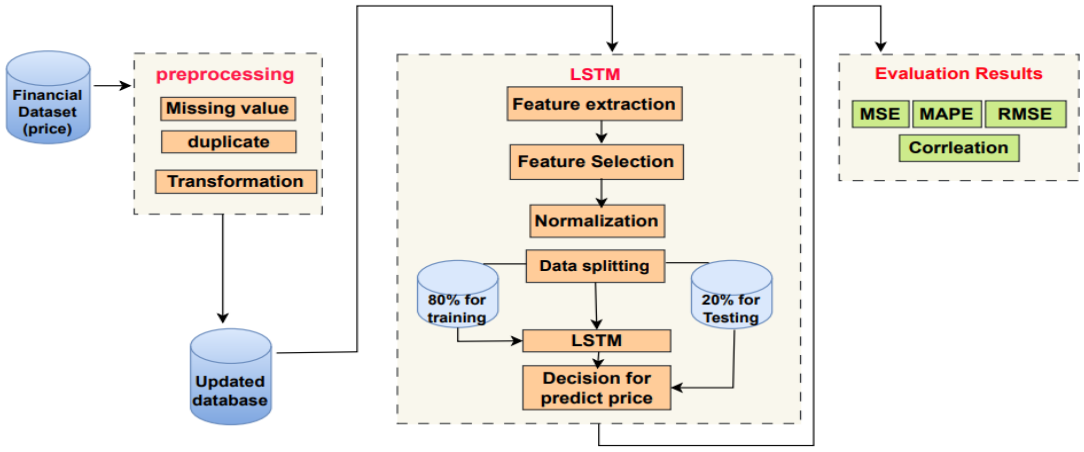

Figure 1. The proposed system architecture

3.1 Dataset

There are numerous global social media platforms. This thesis analyzes the stock exchanges, namely Tesla, S&P500, and NASDAQ, Gold, and News dataset—the study is based on experimental tests with historical data set from global stock markets. The S&P 500 is the second most widely followed market index in the United States. It was established in 1957. The S&P500 comprises 500 prominent firms, including 400 industrial stocks, 40 utility stocks, and 40 others.

There are 20 transportation stocks and several financial stocks [22]. The periods of the S&P500 are selected from 18/11/1999 to 12/07/2022 and include 2417 instances. Tesla is a prominent American automotive firm that manufactures highly advanced electric vehicles. The corporation has recently garnered significant attention due to its stock prices, and the year 2021 witnessed a significant surge in revenue [23][24], rendering TSLA selected from 29/6/2010 to 3/2/2020 to include 5698 instances, as shown visually in figure (2.1) and figure (2.3), respectively. The NASDAQ stock market is a renowned computerized stock exchange in the United States, facilitating the purchase and sale of stocks and shares. It lists around 3,200 businesses, including prominent names like Microsoft, Google, and Intel. Companies such as Amazon and Oracle. It is regarded as the benchmark index for technology stocks in the United States [18]. The period of NASDAQ is selected from 18/11/1999 to 01/04/2020, including 5125 instances, as shown in figure (2.2). The historical data is collected daily (Open et al., Volume, Adj Close), after downloading data set from Kaggel, datasets have saved in tables and they have added to local database. [25].

Table 1: Displays summarization of dataset

DATASET	Period Time	No. of Features	No. of instances
S&p500	18/11/1999 to12/07/2022	7	2417
TESLA	29/6/2010 to 3/2/2020	7	5698
NASDAQ	18/11/1999 to 01/04/2020	7	5125

3.2 Preprocessing stage

Dataset tends to have low quality data. Data processing techniques are classified into two kinds; the first kind is interested in cleaning the data from noisy, missing value, and duplicate data. The second kind concerned on construction the features by TIs generation, reconstructing the data by normalization features, and select the best features and eliminate irrelevant features. These sub-sections describe commonly theories of two kind because they can improve efficiency of the prediction process.

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

3.2.1 Missing values

Ignore the tuple: the strategy of ignoring the tuple with the missing value is used. This means that the record containing the missing value is excluded when performing classification or description tasks. However, this method is not very effective when there is only one record with a missing value. It is also ineffective when the percentage of missing values varies significantly across attributes. In general, there may be more effective methods to handle missing values in training and testing data. [26][27]

3.2.2 Removal of duplicate data: Duplicate entries were eliminated to ensure efficiency and accuracy in the analysis. [28]

3.2.3 Data transformation:

Convert the date format of a dataset including dates to make it appropriate for analysis. One common example of data transformation is converting date formats. Dates can be represented in different formats such as "Month/Day/Year". It may be useful to transform these dates into a standardized format for easier comparison and analysis, such as "Day Month Year", then sorting the data based on dates. Properly sorting the data ensures that events related to the dates are arranged in the correct chronological order.

3.3 Feature extraction

Technical indicators (TIs) are used as inputs to the model to improve the accuracy [6] [29] [20].

Standard deviation (SD) is a mathematical term that quantifies the amount of variation or dispersion around an average value. Standard deviation (SD) is a statistical metric used to quantify volatility. The standard deviation is employed to quantify anticipated risk and ascertain the significance of specific price fluctuations. The calculation of SD is calculated using a particular formula:

Standard Deviation

$$SD = \sqrt{\frac{(Pc(t)-SMA_t(n))^2+(Pc(t-1)-SMA_{t-1}(n))^2+\cdots(Pc(t-n+1)-SMA_{t-n+1}(n))^2}{n}} \quad (1)$$

Where: Pc(t): The value of the variable at time t. SMA_t(n): Simple Moving Average of the variable at time t, calculated by summing the variable from t-n+1 to t and dividing by n.

n: The number of periods in the calculation of the Simple Moving Average.

Relative Strength Index (RSI): Quantifies the velocity and magnitude of price fluctuations [6].

3.3.1 $\text{RSI} = 100 - \frac{100}{1+(\sum_{i=0}^{n-1} \text{UP}_{i-1}/n)/(\sum_{i=0}^{n-1} \text{DW}_{i-1}/n)} \quad (2)$

Where: UP_i-1: The sum of the upward price changes in the previous n periods, divided by n. It represents the average upward price change. DW_i-1: The sum of the downward price changes in the previous n periods, divided by n. It represents the average downward price change. n: The number of periods used in the calculation.

3.3.2 Moving Average Convergence Divergence (MACD): Compares short-term and long-term trends to indicate potential buying or selling opportunities. [30]

3.3.3 MACD Line (MACD) = EMA(Shorter Time Period) - EMA(Longer Time Period) (3)

Signal Line (Signal) = EMA(MACD Line, Signal Time Period) (4)

3 Bollinger Bands: Measures volatility and identifies overbought or oversold conditions.

Standard Deviation: Quantifies the amount of historical volatility in a stock's price.

3.3.4 Price Rate of Change (ROC): Measures the percentage difference in the current closing price and n number of days in the past.[30][6].

ROC is computed by taking the difference between current closing price from n days closing prices ago.

$$\text{ROC}=\frac{P_c(t)-P_c(t-n)}{p_c(t-n)} \quad (5)$$

Where: Pc(t): Current Price is current price of the asset. Pc(t-n): Price n means the past price n periods ago.

3.3.5 Exponential moving average (EMA): is a mathematical method that assigns greater importance to recent data points, resulting in a higher level of significance in the overall computation. This technical indicator is widely popular. The EMA is employed to evaluate the trajectory of a financial asset's movement. To accomplish this goal, EMA utilizes a method known as smoothing, which entails removing the unpredictable variations in price by computing the average price over a designated time, represented as m. The Exponential Moving Average (EMA) is a lagging indicator that relies on historical price data. The EMA lacks the ability to predict upcoming trends, but it is capable of confirming the direction of an already established trend. The Exponential Moving Average (EMA) is a calculation that gives more weight to recent values. It is determined using the following formula [31]:

$$\text{EMA(t)} = (\text{Close(t)} - \text{EMA(t-1)}) * \frac{2}{(\text{n+1})+\text{EMA(t-1)}} \quad (6)$$

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

where EMA[t] is the Exponential Moving Average at time(t). Close[t] is the closing price at time [t]. EMA[t-1] is the Exponential Moving Average at time[t-1]. n is the number of periods

3.3.6 % R (WILLIAMS %R) The current price relationship with the high and low prices over the preceding n days is indicated using the Williams %R indicator. Williams %R is calculated as follows [23]:

$$\%R = \frac{P_H(n) - P_c(t)}{P_H(n) - P_L(n)} \quad (7)$$

Where: P_H(n): The highest high of the security over the previous n periods. P_c(t): The current closing price of the security. P_L(n): The lowest low of the security over the previous n periods.

3.3.7 Commodity Channel Index(CCI) is a technical indicator used in technical analysis to assess the strength of a trend and identify overbought or oversold levels in the market [32][29].

$$\text{CCI} = \frac{P_C(t) - SMAP_C(n)}{(0.015) * Mean\ Deviation} \quad (8)$$

Where: The Typical Price is calculated as the mean of the high, low and close prices which is exactly for a specific time period. The Simple Moving Average (SMA) is the average of the Typical Prices over a specified number of periods. The Mean Deviation is the average of the absolute differences between each Typical Price and the corresponding SMA over the specified number of periods.

3.3.8 Simple Moving Average (SMA): The Simple Moving Average (SMA) can be determined by computing the mean price over a specified time. This operation is performed repeatedly on a daily basis in order to create its own time series. The formula obtained is [33]:

$$SMA = \frac{p_c(t) + p_c(t-1) + \cdots + p_c(t-n-1)}{n} \quad (9)$$

Where: pc(t): The price of the security at time t. n: the number of periods used in calculation.

3.4 Feature selection

The aim of this layer reduces the feature space dimensionality, removes irrelevant features from dataset. It takes the direct effects for application: speedy a DM algorithm, enhancing the data quality and therefore the performance of DM and results are increasing.

3.4.1 Mutual Information MI

MI is used to determine the important features in a dataset. This method relies on the concept of mutual information between the features and the target variable. It calculates the value of mutual information between each feature and the target, and based on that, the features are ranked according to their strength in predicting the target. When using mutual information, importance-ranking values are assigned to each feature between 0 and 1. A value of 0 means that the feature is not important in predicting, while a value of 1 means that the feature is highly important in predicting. By using mutual, the most important features can be easily determined by selecting the top-ranked values in the ranking results. These features can be used in applying a machine learning model to improve performance and save time and resources [34].

The mutual information equation is as follows:

$$MI(X, Y) = \sum \sum P(x, y) * \log(P(x, y) / (P(x) * P(y))) \quad (10)$$

Where: X and Y are random variables for which mutual information is measured. P(x, y) is the joint probability of the occurrence of value x in X and value y in Y. P(x) is the marginal probability of the occurrence of value x in X. P(y) is the marginal probability of the occurrence of value y in Y. log is the natural logarithm.

3.4.2 Principal component analysis (PCA)

PCA, which stands for Principal Component Analysis, is a data analysis technique used for dimensionality reduction. Linear transformation, on the other hand, is an arithmetic process by which information is transformed from one coordinate system to another. For the first variation of any data projection, the value is available in the first coordinate and is referred to as the first principal component. The second variance is also embedded in the second coordinate, also termed the second principal component, and the like. This operation is most often performed in an attempt to reduce the dimension of the given data set while preserving the variation influenced by the attributes contributing best to the data set's variation. This can be achieved by keeping only the first training data vector while discarding the second training data vector which is the principal component. In aggregate lower-order components we may preserve our self-most essential features of initial data. The input is a collection of vectors, which are m-by-n matrices containing samples. [7] [35]

- Computing the mean: The mean is computed for each variable by:

$$\mu = \left(\frac{1}{n}\right) * \sum(X_i) \quad (11)$$

- Normalizing the data: The data is normalized by subtracting the mean from each value:

$$x_i = x_i - \mu \quad (12)$$

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

- Computing the covariance matrix: The covariance matrix is computed, which contains the correlation between all possible pairs of variables:

$$\Sigma = \frac{1}{n} * \sum Xi * X_i^T \quad (13)$$

- Computing the eigenvalues and eigenvectors:

Calculate the eigenvalues (λ) and eigenvectors (V) of the covariance matrix Σ.
Computing the principal components:
Multiply the normalized data by the eigenvectors to obtain the principal components:

$$Y_i = X_i * V \quad (14)$$

Where: xi is the original variable, μ is the mean, xi' is the normalized value (shifted variable)
, Σ is the covariance matrix, n is the number of samples; V is the matrix of eigenvectors.

3.5 Min-Max normalization

The input data is transformed via linear scaling to a range of values between 0 and 1, resulting in a normalized multi-dimensional time-series. The act of scaling this data enhances the efficiency of the training method for the neural network model. The value of x0 can be stated using linear scaling.[36]:

$$\upsilon^* = \frac{(\upsilon - \mu in)}{(\mu ax - \mu in)} \quad (15)$$

Where :υ *: is the new value ; υ: is the old value , μin: the minimum value The data has been normalized using a min–max scaler , μax: the maximum value.

3.6 Data splitting

The correct measurement of data mining techniques demands the use of test data that were not yet intended during the training stage previously. Through Cross Validation (CV), a section of the SM dataset is divided into two segments on every fold. The data is split into two subsets: the training set is 80% of the data and here the validation set is 20% of the training data.

3.7 LSTM model description

Long short-term memory is a type of recurrent neural network. Though RNNs work well on time series data, it is difficult to avoid the problem of long-term dependency due to the vanishing gradient [37]. When you first look at the engineering of LSTM figure (5), it seems more difficult. The key distinction is that it has three outputs rather than two like standard recurrent neural networks. The first output is the memory matrix, which has not been transformed with non-linear functions, making it considerably easier to optimize and solve the vanishing gradient problem. The second and third outputs are same, and they are sent as a conventional output from the layer to the next input cell. A single cell is partitioned into three components: the forget gate, the input gate, and the output gate. The forget gate is a sigmoid function that receives the current input (xt) and the previous output (ht-1) [22][38]. This function generates a matrix that has values between 0 and 1. The resulting matrix is then multiplied by the memory matrix. This process can be referred to as a gate, where values that are close to 0 are discarded in the memory matrix, while values that are close to 1 are preserved. The input gate incorporates additional weights into the memory matrix. The sigmoid function determines the impact of each individual input data point, which is then subjected to a non-linear transformation using the tanh function. The weight matrix generated in this manner is incorporated into the memory cell. The output gate capitalizes on the previous work done to construct the memory cell. Unlike ordinary neural networks, a single layer does not only modify the input, but it is also fine-tuned by the memory matrix. This matrix preserves important data about major relationships from previous instances. [23][39].

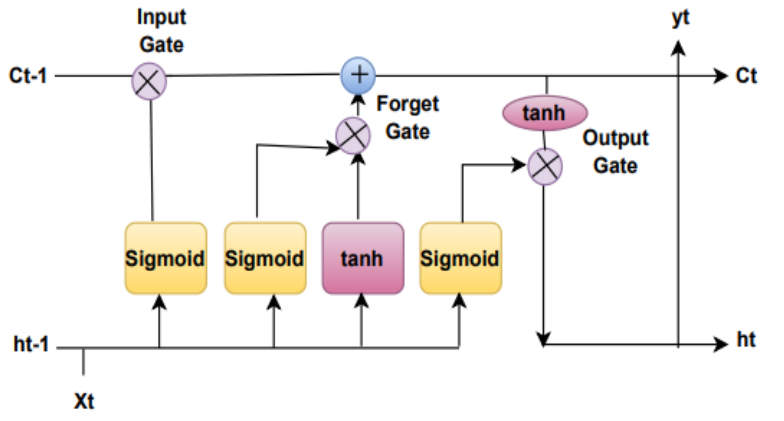

Figure 2. The internal architecture of a LSTM

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

Fig 2 shows a pictorial representation of LSTM gates and their mechanism in this context:

Forget Gate:

Previous Cell State (Ct−1): It represents the state of the previous cell.
Forget Gate (ft): It is a sigmoid layer that decides which information should be forgotten in the current cell.
Output from the Forget Gate (ft * Ct−1): This output is added to the current cell (Ct).
“

$$f_t = \sigma(X_t * U_t + H_{t-1} * w_t) \quad (16)$$

Where: X_t: input to the current timestamp, U_f: weight associated with the input, Ht-1: the hidden state of the previous timestamp, W_f: it is the weight matrix associated with the hidden state”,
Input Gate:
Input Signal (X_t): It is the sequential input signal at the current time step.
Input Gate (it): It is a sigmoid layer that determines what information should be updated in the current cell.
Updated Values (C'): These are the values computed by a *tanh* layer.
Output from the Input Gate (it * C'): This output is added to the current cell (Ct) after the gate's decision.

$$i_t = \sigma(X_t * U_t + H_{t-1} * w_t) \quad (17)$$
$$N_t = tanh(X_t * U_t + H_{t-1} * w_t) \quad (18)$$
$$C_t = (f_t * C_{t-1} + i_{t-1} * N_t) \quad (19)$$

Where X_t: input at the current timestamp t, U_i: weight matrix of input, H_{t-1}: A hidden state at the previous timestamp, W_i: weight matrix of input associated with hidden state. C_{t-1} is the cell state at the current timestamp.

Output Gate:

Current Cell (*Ct*): It represents the current cell state.
Output Gate (*Ot*): It is a sigmoid layer that determines the output from the current cell.
Final Output (*Ht*): This output is computed by element-wise multiplication between the Output Gate (*Ot*) and the tanh value of the current cell (Ct).

$$Ot = \sigma(X_t * U_o + H_{t-1} * w_o) \quad (20)$$
$$H_t = O_t * \tanh(C_t) \quad (21)$$

3.7.1 Design the LSTM Neural Network

This model is a simple LSTM model used for time series prediction, where it is trained using the training data and Adam optimization algorithm with the mean squared error loss function.

LSTM Layer:

This layer is used to handle sequential data. It is one of the main layers in deep neural networks and helps in understanding the temporal relationships and dependencies between points in the time series. It use number of units is 128 , activation is 'tanh'.

$$f(x) = \tanh x = \frac{\sinh x}{\cosh x} = \frac{e^a - e^{-a}}{e^a + e^{-a}} \quad (22)$$

Second LSTM layer: Used to create an LSTM layer with 64 units. It uses the "tanh" activation function to activate the output from the LSTM units. This layer is used to aggregate and summarize information from the first LSTM layer.

Dropout layer: The dropout layer with a rate of (0.3). Dropout is added to prevent overfitting, reduce overfitting, and increase generalization. This helps to avoid over-reliance on specific units and increases the model's ability to generalize.

Dense layer: Used to create a fully connected layer with 50 units. Data passes between consecutive layers and helps to transform the information into predictive representations.

Dense Layer: This layer is used to transform the outputs from the previous layers into final predictions. It uses a single unit to predict a single value.

Model Compilation: In this step, the algorithm used for weight updates is specified, and the model is prepared for training. The Adam optimizer is used for weight updates. The Mean Squared Error loss function is used to measure the difference between the expected and actual values.

Model Training: In this step, the model is trained on the training data. The number of training epochs is (100) and the batch size (32) used in each weight update are specified. The progress bar is displayed during training.

Optimizer: The optimization process is about finding the best parameters and values for the kernels and biases, to set the training operation correctly (as described in the chapter two). In our approach, the Adam optimizer is chosen. Due to its high performance. The number of epochs is set to 100; the batch size is set to 32.

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

4. Evaluating model

Accurate assessment of data mining techniques requires the use of test data that has not been previously seen during the training phase. In Cross Validation (CV), the SM dataset is divided into two sections in each fold. 80% of the data is allocated as the training set, whereas 20% of the training data is selected as the validation set. In order to assess the efficacy of the model. applied on three distinct sector businesses, including Tesla, NASDAQ, and S&P500, utilizing the Attention model, we utilize significant machine learning and deep learning algorithms to assess their performance using evaluation metrics such as: [6][40][2] [32]

1**. Mean squared error (MSE**): It is a measure that gives the amount by which an actual value differs from the expected value. It is arrived at by subtracting one value from another, squaring the result and then averaging the two. A lower MSE score implies that it is more accurate in prediction as compared to other sets of data inputs. As the following equation:

$$MSE = \frac{\sum(\boldsymbol{xi}-\boldsymbol{xbi})^{\boldsymbol{2}}}{\boldsymbol{n}} \quad (23)$$

Where xi: The actual value of the target variable at index I. x̂i: The predicted value of the target variable at index i. n: The total number of samples or data points.

2. Root Mean Squared Error (RMSE): is gained in the form of the square root of the average of the squared deviations of the obtained and expected values. It can be often used as the metric for model evaluation in the models of regression, and the smaller value of RMSE is considered better. Its equation as the following:

$$\text{RMSE}=\sqrt{\frac{\sum_i^n(yi-ybi)^2}{n}} \quad (24)$$

Where yi: The actual value of the target variable at index I, ŷi: The predicted value of the target variable at index i , n: The total number of samples or data points.

3. Mean Absolute Error (MAPE):

It is the mathematical average of taken differences between forecasted and real amounts. MAE applies the concept of average absolute error that sums up discrepancies between forecasted and actual values. MAPE measure is often used when comparing regression models since it indicates the average degree of error in measurements. While using MAPE there is no such penalty for large mistakes as is the case with RMSE, which makes MAE less sensitive to outliers. Its equation:

$$\text{MAPE}=\frac{\sum_i^n\left|\frac{(yi-ybi)}{yi}\right|}{n}*100 \quad (25)$$

Where: yi: The actual value of the target variable at index I, ŷi: The predicted value of the target variable at index i, n: The total number of samples or data points.

4. Correlation coefficient: calculates the correlation between actual and anticipated values.

The estimated correlation coefficient measures the strength of the association between the variables.

5. Results and discussion

This paper of stock price prediction based on the PCA-LSTM model is proposed conducting an experiment using three datasets of stock indexes: TESLA, NASDAQ, and S&P 500. Initially, mutual information is employed in the data, and the important features are selected. The feature names and their corresponding scores are obtained, and the features are ranked based on these scores. Each feature is assigned a weight. A threshold is set to retain only the major components with weights above the threshold. The total count of the selected components is calculated. Finally, the selected components are derived from the original dataset. Then, PCA was used to select features with a decrease in features (17 TIs and 7 standard features) because each feature adds another dimension to the search space. By reducing the number of irrelevant features, the number of principal components is determined by the previously selected important features. The work for the predicated can be simplified. Then, the LSTM model is implemented using feature selection. In the comparison of LSTM with other approaches in stock index prediction, the proposed system achieves superior performance with a LOSS of 0.001. It consistently achieves a CORRELATION of 99% on the stock indexes dataset. This demonstrates the effectiveness of our technique in analyzing prediction jobs on financial time series. Irrespective of the type of dataset. The experimental findings have demonstrated that the proposed model attained a remarkable CORRELATION of 99% and an MSE 0.0001, MAE 0.009, and RMSE 0.01 during epoch 100 for the Tesla dataset, as shown in Figure 3. Furthermore, the model exhibited great performance when applied to the S&P500 and NASDAQ datasets, as illustrated in Figures (4) and (5), respectively.

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

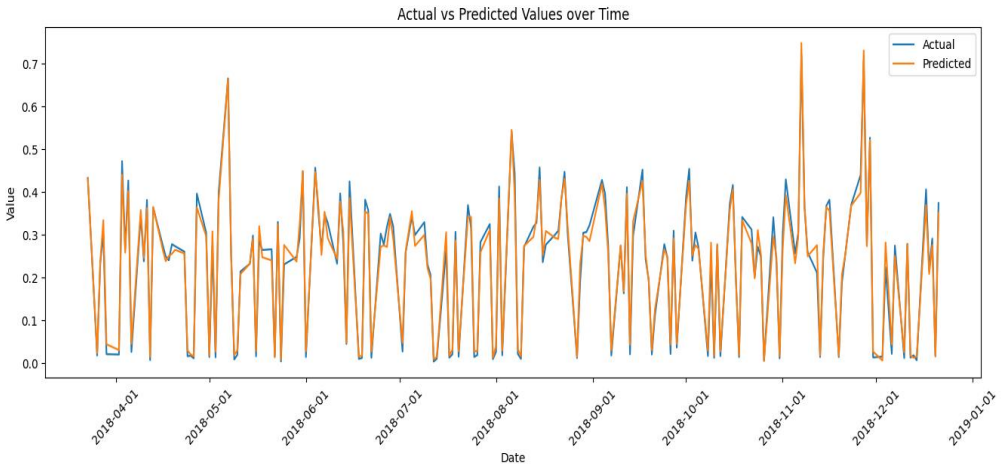

Figure 3. The actual and predicted values of the price in Tesla Dataset with LSTM

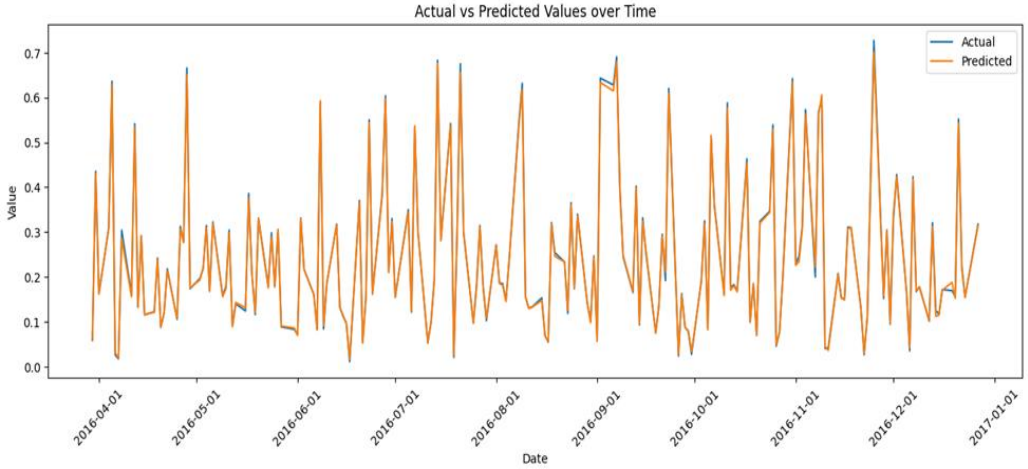

Figure 4. The actual and predicted values of the price in NASDAQ Dataset with LSTM

FIGER (4) shows the Predicted closing price (orange line) of NASDAQ stocks vs. the actual closing value (Blue line) values of the price over time. The error between the actual and predicted values is calculated. The dates from the original dataset are used. The proposed model attained a remarkable Nasdaq database CORRELATION of 99% and an MSE 0.0001, MAE 0.001, RMSE 0.01 during epoch 100.

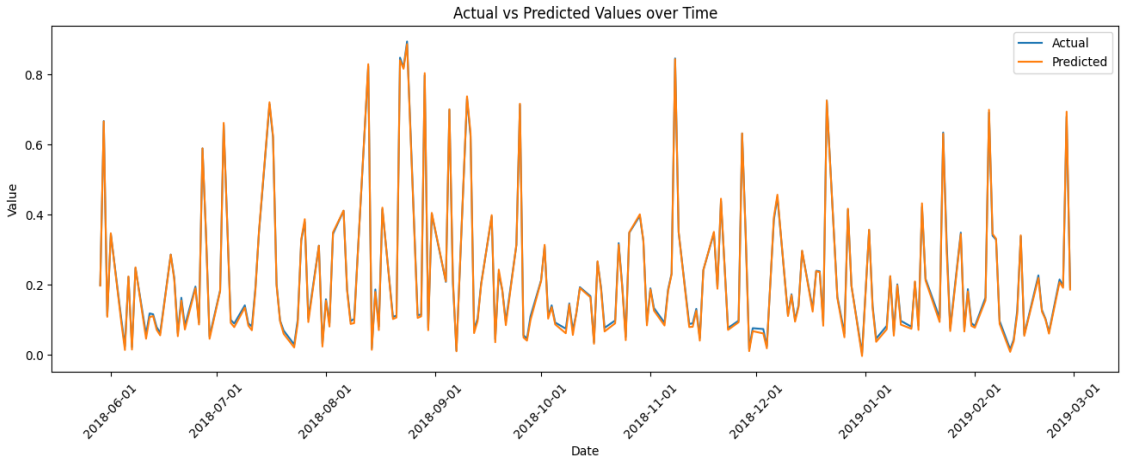

Figure 5. The actual and predicted values of the price in S&P500 Dataset with LSTM

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

Figure (5) shows the Predicted closing price (orange line) of S&P500 stocks vs. the actual closing value (Blue line) values of the price over time. The error between the actual and predicted values is calculated. The dates from the original dataset are used. The proposed model attained a remarkable s&p500 database CORRELATION of 99% and an MSE 0.0003, MAE 0.01, and RMSE 0.01 during epoch 100. To verify the comparison effect of the proposed system in this study with other methods used in the past, the system was compared with a deep learning prediction model utilizing MUTUAL, PCA, and LSTM methods. The comparison results between the proposed system and other methods shown in Table (2) are derived, displaying the techniques used in the various research papers and their corresponding results. The table also shows the results of the proposed prediction system.

Table 2: Comparative analysis of previous and proposed study

Ref	Method	Dataset	Evaluation metrics
[12]	LSTM	National Stock Exchange (NSE)	MAPE<50
[13]	LSTM	Tehran stock exchange.	MAPE=0.1369, MAE=0.100
[14]	LSTM	banks (AXIS, HDFC, ICICI, KOTAK and SBI) 6.1.3	RMSE=3.15
[15]	LSTM	Tesla Inc	MSE=0.69205, RMSE=0.8318
[16]	PCA_LSTM	stock information of Pingan Bank comes from RESSET financial research platform	MSE=0.221, MPE=1.667
[17]	LSTM	S&P500 NASDAQ	MAPE=0.75 MAPE=0.84
[18]	LSTM	NASDAQ S&P500	RMSE for s&p500 =34.86 RMSE for NASDAQ=26.93 MAPE for s&p500 =64.2600 MAPE for NASDAQ=56.19
[19]	P-FTD_LSTM	S&P500, SSE, and KOSPI	MAPE 0.325, MAE 7.795, RMSE 10.669
[20]	RNN	S&P500 companies	MAPE=2.03
[21]	DECISION TREE	Tesla Inc	RMSE=9267.3
Proposed system	MUTUAL Information, PCA, LSTM	NASDAQ	MSE=0.0001, MAE=0.009, RMSE=0.01, CORRELATION=0.99, LOSS=0.002
		S&P500	MSE=0.0003, MAE=0.01, RMSE=0.01, CORRELATION=0.99, LOSS=0.001
		TSLA	MSE=0.0001, MAE=0.01, RMSE=0.01, CORRELATION=0.99, LOSS=0.001

6. Conclusion

In this study, the technique indicators and the MUTUAL information approach are used and then followed by the principal component analysis to decrease the number of factors influencing the stock prices. To enhance the accuracy of the predictions made, deep learning is applied where LSTM with Adam optimizer and Tanh activation function is used. The assessment metrics include Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The experiments are conducted on S&P 500 index, TSLA stock, and NASDAQ index. According to the results of the experiment, applying the given approach to model the forecast of the given data sample will allow obtaining the improved results of the forecast in comparison with the majority of the similar works. In the foreseen future, the application of learning and mastering the market through the help of machine learning techniques will result to a promising future.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

References

[1] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "ScienceDirect Stock Closing Closing Price Price Prediction Prediction using using Machine Machine Learning Learning Techniques Techniques," Procedia Comput. Sci., vol. 167, no. 2019, pp. 599–606, 2020, doi: 10.1016/j.procs.2020.03.326.

[2] D. K. Padhi, N. Padhy, A. K. Bhoi, J. Shafi, and S. H. Yesuf, "An Intelligent Fusion Model with Portfolio Selection and Machine Learning for Stock Market Prediction," Comput. Intell. Neurosci., vol. 2022, 2022, doi: 10.1155/2022/7588303.

[3] P. Lee, Z. Huang, and Y. Tang, "Trend Prediction Model of Asian Stock Market Volatility Dynamic Relationship Based on Machine Learning," Secur. Commun. Networks, vol. 2022, 2022, doi: 10.1155/2022/5972698.

[4] X. Zhou, Z. Pan, G. Hu, S. Tang, and C. Zhao, "Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets," Math. Probl. Eng., vol. 2018, 2018, doi: 10.1155/2018/4907423.

[5] M. C. Wu, S. Y. Lin, and C. H. Lin, "An effective application of decision tree to stock trading," Expert Syst. Appl., vol. 31, no. 2, pp. 270–274, 2006, doi: 10.1016/j.eswa.2005.09.026.

[6] Y. Lin, S. Liu, H. Yang, and H. Wu, "Stock Trend Prediction Using Candlestick Charting and Ensemble Machine Learning Techniques with a Novelty Feature Engineering Scheme," IEEE Access, vol. 9, pp. 101433–101446, 2021, doi: 10.1109/ACCESS.2021.3096825.

[7] C. Lohrmann and P. Luukka, "Classification of intraday S&P500 returns with a Random Forest," Int. J. Forecast., vol. 35, no. 1, pp. 390–407, 2019, doi: 10.1016/j.ijforecast.2018.08.004.

[8] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," Phys. D Nonlinear Phenom., vol. 404, no. March, pp. 1–43, 2020, doi: 10.1016/j.physd.2019.132306.

[9] H. Boubaker, B. Saidane, and M. Ben Saad Zorgati, "Modelling the dynamics of stock market in the gulf cooperation council countries: evidence on persistence to shocks," Financ. Innov., vol. 8, no. 1, 2022, doi: 10.1186/s40854-022-00348-3.

[10] J. Cao, "Stock price forecasting model based on modified convolution neural network and financial time series analysis," no. November 2018, pp. 1–13, 2019, doi: 10.1002/dac.3987.

[11] M. Göçken, M. Özçalıcı, A. Boru, and A. T. Dosdoğru, "Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection," Neural Comput. Appl., vol. 31, no. 2, pp. 577–592, 2019, doi: 10.1007/s00521-017-3089-2.

[12] M. Iyyappan, S. Ahmad, S. Jha, A. Alam, M. Yaseen, and H. A. M. Abdeljaber, "A Novel AI-Based Stock Market Prediction Using Machine Learning Algorithm," Sci. Program., vol. 2022, 2022, doi: 10.1155/2022/4808088.

[13] O. Kelany, S. Aly, and M. A. Ismail, "Deep Learning Model for Financial Time Series Prediction," Proc. 2020 14th Int. Conf. Innov. Inf. Technol. IIT 2020, pp. 120–125, 2020, doi: 10.1109/IIT50501.2020.9299063.

[14] M. Mazed, "Stock Price Prediction Using Time Series Data," Brac Univ., vol. 1(1), no. August, pp. 1–51, 2019.

[15] K. Patel and S. Saket, "Analysis of Stock Market Forecasting using Machine Learning Mechanism," JASC J. Appl. Sci. Comput., vol. V, no. 112, pp. 112–119, 2018.

[16] Y. Wen, P. Lin, and X. Nie, "Research of stock price prediction based on PCA-LSTM model," IOP Conf. Ser. Mater. Sci. Eng., vol. 790, no. 1, 2020, doi: 10.1088/1757-899X/790/1/012109.

[17] G. Bathla, "Stock price prediction using LSTM and SVR," PDGC 2020 - 2020 6th Int. Conf. Parallel, Distrib. Grid Comput., pp. 211–214, 2020, doi: 10.1109/PDGC50313.2020.9315800.

[18] U. Gurav and N. Sidnal, "Adaptive Stock Forecasting Model using Modified Backpropagation Neural Network (MBNN)," Proc. Int. Conf. Comput. Tech. Electron. Mech. Syst. CTEMS 2018, pp. 380–385, 2018, doi: 10.1109/CTEMS.2018.8769290.

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

[19] D. Song, A. M. Chung Baek, and N. Kim, "Forecasting Stock Market Indices Using Padding-Based Fourier Transform Denoising and Time Series Deep Learning Models," IEEE Access, vol. 9, pp. 83786–83796, 2021, doi: 10.1109/ACCESS.2021.3086537.

[20] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," Proc. - 5th IEEE Int. Conf. Big Data Serv. Appl. BigDataService 2019, Work. Big Data Water Resour. Environ. Hydraul. Eng. Work. Medical, Heal. Using Big Data Technol., pp. 205–208, 2019, doi: 10.1109/BigDataService.2019.00035.

[21] S. E. Arefin, "Second Hand Price Prediction for Tesla Vehicles," 2021, [Online]. Available: http://arxiv.org/abs/2101.03788

[22] M. Kijewski and R. Ślepaczuk, "Predicting Prices of S&P500 Index Using Classical Methods and Recurrent Neural Networks," vol. 2020, no. 27, 2020.

[23] T. H. H. Aldhyani and A. Alzahrani, "Framework for Predicting and Modeling Stock Market Prices Based on Deep Learning Algorithms," Electron., vol. 11, no. 19, 2022, doi: 10.3390/electronics11193149.

[24] A. H. Khan et al., "A performance comparison of machine learning models for stock market prediction with novel investment strategy," PLoS One, vol. 18, no. 9 September, pp. 1–19, 2023, doi: 10.1371/journal.pone.0286362.

[25] G. Edman and M. Weishaupt, "Predicting Tesla Stock Return Using Twitter Data," no. May 2020, 2020.

[26] S. Lundgren, "1. Introduction 11," Fight Against Idols, 2016, doi: 10.3726/978-3-653-01927-8/2.

[27] M. Biswas, A. Shome, M. A. Islam, A. J. Nova, and S. Ahmed, "Predicting stock market price: A logical strategy using deep learning," ISCAIE 2021 - IEEE 11th Symp. Comput. Appl. Ind. Electron., pp. 218–223, 2021, doi: 10.1109/ISCAIE51753.2021.9431817.

[28] O. Rivera, "Enhancing Influencer Marketing Strategies through Machine Learning: Predictive Analysis of Influencer-Generated Interactions," vol. TRITA-EECS, no. 2023:0000. KTH Royal Institute of Technology, 2023. [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1783645%0Ahttps://www.diva-portal.org/smash/get/diva2:1783645/FULLTEXT01.pdf

[29] Z. Xie and Y. Wang, "Exploration of Stock Portfolio Investment Construction Using Deep Learning Neural Network," Comput. Intell. Neurosci., vol. 2022, 2022, doi: 10.1155/2022/7957097.

[30] K. J. Kim, "Financial time series forecasting using support vector machines," Neurocomputing, vol. 55, no. 1–2, pp. 307–319, 2003, doi: 10.1016/S0925-2312(03)00372-2.

[31] J. Ayala, M. García-Torres, J. L. V. Noguera, F. Gómez-Vela, and F. Divina, "Technical analysis strategy optimization using a machine learning approach in stock market indices[Formula presented]," Knowledge-Based Syst., vol. 225, 2021, doi: 10.1016/j.knosys.2021.107119.

[32] H. Chung and K. shik Shin, "Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction," Neural Comput. Appl., vol. 32, no. 12, pp. 7897–7914, 2020, doi: 10.1007/s00521-019-04236-3.

[33] E. El., "A New Particle Swarm Optimization Based Stock Market Prediction Technique," Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 4, pp. 322–327, 2016, doi: 10.14569/ijacsa.2016.070442.

[34] H. Wu, A. Gattami, and M. Flierl, "Conditional mutual information-based contrastive loss for financial time series forecasting," ICAIF 2020 - 1st ACM Int. Conf. AI Financ., 2020, doi: 10.1145/3383455.3422550.

[35] L. Hang, D. Liu, and F. Xie, "A Hybrid Model Using PCA and BP Neural Network for Time Series Prediction in Chinese Stock Market with TOPSIS Analysis," Sci. Program., vol. 2023, 2023, doi: 10.1155/2023/9963940.

[36] H. Chung and K. Shin, "Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction," Neural Comput. Appl., vol. 32, pp. 7897–7914, 2020.

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024

[37] T. J. Prins, "UCLA UCLA Electronic Theses and Dissertations Title," 2019, [Online]. Available: https://escholarship.org/uc/item/0th2s0ss

[38] H. Kiran, S. Surayagari, A. Ben-Hur, and C. Stein, "THESIS STOCK MARKET PREDICTIONS USING MACHINE LEARNING Submitted by," 2021.

[39] P. Chhajer, M. Shah, and A. Kshirsagar, "The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction," Decis. Anal. J., vol. 2, no. November 2021, p. 100015, 2022, doi: 10.1016/j.dajour.2021.100015.

[40] S. Albahli, A. Awan, T. Nazir, A. Irtaza, A. Alkhalifah, and W. Albattah, "A deep learning method DCWR with HANet for stock market prediction using news articles," Complex Intell. Syst., vol. 8, no. 3, pp. 2471–2487, 2022, doi: 10.1007/s40747-022-00658-0.

DOI: https://doi.org/10.54216/FPA.170114
Received: November 25, 2023 Revised: March 17, 2024 Accepted: July 18, 2024