Grenze International Journal of Engineering and Technology, June Issue

Time Series Analysis on Stock Market and Forecasting

Krishna Yadav[1], Ajay Pal Singh[2], Rohit Sharma[3] and Rakesh Chouhan[4]
[1-4]Department of computer Science & Engineering, Chandigarh University, Punjab, India
Email: 19bcs3034@gmail.com, apsingh3289@gmail.com, 19bcs2700@gmail.com, 19bcs2987@cuchd.in

***Abstract*— Forecasting stock market trends is challenging due to the complexity and dynamics of the system. Time series analysis has become a popular method in recent years for predicting such trends. This research paper examines the use of time series analysis in stock market forecasting and evaluates the effectiveness of different statistical techniques in predicting future market trends. It examines multiple time series models, such as ARIMA and LSTM, to assess their performance in predicting the stock market. Additionally, the study explores the influence of economic indicators on stock prices and their incorporation into time series models to improve forecasting accuracy. The results suggest that the combination of time series analysis and economic indicators can be a useful approach to forecast stock market trends. However, the precision of the forecasts relies on the quality of historical data and specific market conditions. The study emphasizes the importance of acknowledging the uncertainty involved in stock market forecasts and the need for regular monitoring and adjustment of forecasting models. This research paper also provides significant insights into the practical application of time series analysis in forecasting stock market trends and underscores its potential in enhancing investment strategies. By leveraging historical data to identify significant patterns and trends, investors and analysts can gain deeper insights into future market performance, making well-informed investment decisions that lead to more successful outcomes.**

***Index Terms*— forecasting, stock market, time series, ARIMA, LSTM, Stock prices, indicators, patterns and trends**

I. Introduction

For investors, traders, and financial analysts, the prediction of stock market trends has consistently been a crucial domain of study and examination. Accurate predictions are crucial for investors to make informed decisions about when to buy or sell stocks, whereas traders employ forecasting techniques to profit from short-term market fluctuations. Financial analysts, on the other hand, use forecasting models to assess the health of individual companies or industries and to identify potential investment opportunities. In this context, time series analysis is a vital tool for stock market forecasting, as it provides a systematic framework for analyzing the complex patterns and trends that characterize financial data. By analyzing historical stock prices, time series analysis can help identify patterns such as seasonality, trends, and cycles, which can then be used to make accurate predictions about future prices. As a barometer of economic health and a significant driver of growth, the stock market plays a pivotal role in the global economy. For investors and financial analysts, the ability to accurately predict stock market trends is essential for making informed investment decisions. However, forecasting the behavior of the stock market is a challenging task due to its inherent volatility and unpredictability.
Time series analysis has become increasingly important in stock market forecasting as it provides a systematic

Grenze ID: 01.GIJET.9.2.71

analyzing the complex patterns and trends that characterize financial data. This approach involves examining historical data to identify patterns and trends, such as seasonality and cycles, which can be used to make accurate predictions about future market trends. Time series analysis has gained popularity in recent years due to its ability to capture the inherent volatility of the stock market and its flexibility in adapting to changes in market conditions.
One approach that has gained popularity in recent years is time series analysis encompasses the examination of historical data to detect patterns and trends in stock market information throughout a period. Time series analysis can be used to predict future market trends by identifying and modeling the underlying patterns in the data. This method allows investors and analysts to make more informed decisions based on past market trends, and to identify key indicators that may signal potential changes in the market.
In recent years, the employment of time series analysis for stock market prediction has gained momentum in recent years. This approach has become increasingly popular due to its ability to capture the inherent volatility of the stock market and its flexibility in adapting to changes in market conditions. Additionally, time series analysis can incorporate a wide range of economic and financial data, allowing analysts to consider a variety of factors that may influence market trends.
The goal of this research paper is to investigate how time series analysis can be applied in predicting stock market trends, providing an overview of the statistical techniques used and their effectiveness in predicting future market trends. The paper will compare the performance of different time series models, including ARIMA models and LSTM Model, and will investigate the impact of various economic indicators on stock prices.
The research will also discuss the potential benefits and limitations of using time series analysis in stock market forecasting and its implications for investors and financial analysts. Specifically, the paper will explore how time series analysis can be used to improve investment strategies and to identify potential risks and opportunities in the stock market.
Overall, this research paper contributes to our understanding of the role of time series analysis in stock market forecasting, providing valuable insights for investors and analysts seeking to improve their investment strategies. By using historical data and identifying key patterns and trends, investors and analysts can make more informed decisions about future market performance, ultimately leading to more successful investment outcomes.

II. Literature Review

The proposed framework for stock price forecasting in this study utilizes stacked autoencoders and long-short term memory (LSTM) networks in a deep learning approach. By applying these methods for feature extraction and temporal dependency modelling, the research focuses on increasing the accuracy of stock price prediction. Using real financial time series data, the suggested model is evaluated, and the forecast accuracy yields encouraging results [1].
An attention-based LSTM model for forecasting stock prices. Goals are to make the model easier to understand and to pinpoint the important variables that influence stock price movements. Using a sizable dataset of stock market data, the suggested model is assessed, and it shows promising results in terms of prediction accuracy. The attention mechanism offers insights into the dynamics of the stock market and aids in the identification of the major variables that affect stock price movements [2,3].
A multi-task learning framework for concurrently forecasting stock prices and trading volumes. Focuses on how news sentiment affects prediction accuracy and looks at how well sentiment analysis can be included into the prediction model. A substantial dataset of financial news and stock market data is used to assess the proposed framework, and the findings show that include news sentiment can increase the precision of forecasts for both stock price and trading volume. Sheds light on the advantages that can result from include non-financial variables in stock market forecasting algorithms [4,5].
The effectiveness of several time series analysis and machine learning methods for forecasting stock prices is examined. The accuracy of several models, including ARIMA, exponential smoothing, and artificial neural networks, is compared in the study. Evaluation of the effects of various elements, such as the selection of input variables and the volume of training data, on model performance is the main objective. Findings shed light on the relative effectiveness of various models and their applicability to various types of data [6].
Using various technical indicators as features in a deep belief network (DBN) to forecast stock prices and focuses on how feature selection affects prediction accuracy and assesses how well various technical indicators can forecast stock prices. A substantial dataset of stock market data is used to evaluate the suggested model, and the findings indicate that using a variety of technical indicators can increase the precision of stock price forecasts and the possible advantages of utilizing DBN and technical indicators to forecast the stock market [7].

The time series are broken down into various frequency components using wavelet transform, and then ARIMA and LSTM models are applied to the resultant sub-series. This work presents a hybrid strategy combining wavelet transform, ARIMA, and LSTM for stock price prediction. As compared to standalone ARIMA and LSTM models, the technique is proven to improve model performance [8].

In this study, the effectiveness of LSTM and other machine learning and deep learning algorithms for forecasting stock price movements is compared. The study uses a sizable dataset of stock market data to assess the accuracy of several models. The emphasis is on evaluating the accuracy and stability of each model's performance, with a special emphasis on how the duration of the training data affects prediction accuracy. The study's findings shed light on the relative advantages and disadvantages of various models as well as their applicability to various kinds of stock market data [9, 10].

A hybrid model that combines the grey wolf optimizer (GWO) and LSTM. The goal is to decrease forecasting errors by using the GWO algorithm to optimize the LSTM model's hyper parameters. The effectiveness of the hybrid model is evaluated using a substantial stock market dataset, and the results indicate that it surpasses conventional LSTM models regarding the accuracy of stock price predictions. The potential of employing GWO as an optimization tool to boost the precision of LSTM-based stock market prediction models [11].

The LSTM and XGBoost algorithms is to increase forecast stability and accuracy by taking advantage of each model's advantages. The findings show that the proposed ensemble model may achieve greater prediction accuracy and stability compared to individual LSTM and XGBoost models [12].

With an emphasis on precision and stability a deep learning-based model for predicting the stock price index. In order to identify both short-term and long-term patterns in the stock market data, the study creates a hybrid model that combines a convolutional neural network (CNN) with an LSTM network. The proposed model is tested using a large dataset of stock market data, and the results demonstrate that it outperforms traditional time series models in both accuracy and stability [13, 14].

III. METHEDOLOGY USED FOR ANALYSIS

Several methodologies have shown that Machine Learning and Deep Learning can benefit from Time-Series Analysis models. The present project builds upon this by utilizing the ARIME and LSTM models, which are renowned for their forecasting capabilities and serve as the cornerstone of the proposed technique.

There are several processes involved in the approach for creating ARIMA and LSTM models. Starting with the collection of historical time series data and guaranteeing stationarity, ARIMA prepares its data. Following model identification, inspection of the ACF and PACF plots is used to establish the order of the differencing (d), autoregressive (p), and moving average (q) components. After fitting it to the training data and making any required modifications, the model is then estimated. The performance of the model is then assessed by making predictions for future values and contrasting them with the actual values from the validation set using measures like MSE or MAE. Adjustments can be made or the procedure repeated if necessary.

Data preparation for LSTM models include obtaining and scaling historical time series data. The LSTM model architecture is then established after dividing the data into training and validation sets. Iteratively feeding input sequences and modifying parameter values using optimization techniques like gradient descent are both involved in training the model. The validation set is used to assess the model's performance, and new input sequences are fed into the trained model to conduct forecasting. Accuracy is determined by comparing the predicted values to the actual values and using assessment measures. The model architecture or training procedure can be modified as necessary.

A. Auto-Regressive Integrated Moving Average (ARIMA)

The ARIMA model explains the relationship between consecutive observations by describing how each subsequent observation is linked to the preceding one.

The formula for the ARIMA model can be written as:

$$Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + ... + \Phi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad \text{…………(i)}$$

where:

- Y_t: value of the time series at time t
- c: constant or intercept term
- $\Phi_1, \Phi_2, ..., \Phi_p$: autoregressive coefficients of order p
- ε_t: : the error term at time t, assumed to be white noise with zero mean and constant variance
- $\theta_1, \theta_2, ..., \theta_q$: moving average coefficients of order q

The ARIMA model is frequently employed in conjunction with other time series forecasting methods, including seasonal decomposition, trend analysis, and exponential smoothing.
Time series data may contain both a connection with current lags and a seasonal periodic component, with every s observations repeating. For example, a monthly time series of observations may be influenced by yearly lags such as Xt-12, Xt-24, and so on. The ARIMA model is used to address this situation.

ARIMA(p, d, q)(P, D, Q)m…………(ii)

where:
- p: the number of autoregressive terms (AR)
- d: the degree of differencing (I)
- q: the number of moving average terms (MA)
- P: the number of seasonal autoregressive terms (SAR)
- D: the degree of seasonal differencing (SI)
- Q: the number of seasonal moving average terms (SMA)
- m: the number of periods in each season

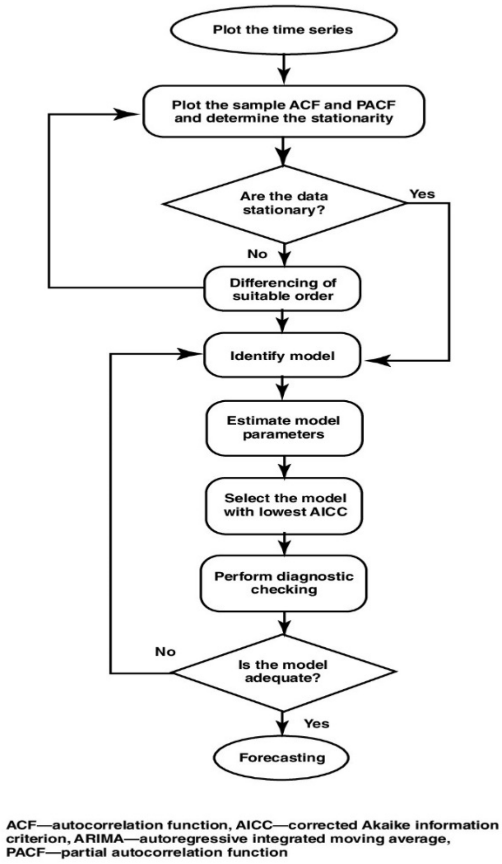

Fig. 1. Modelling Process of ARIMA

B. Long Short-Term Memory (LSTM)

In order to overcome the limitations of traditional RNNs in modeling long-term dependencies in sequential data, Recurrent Neural Networks (RNNs) such as Long Short-Term Memory (LSTM) networks are developed. LSTMs are highly suited for modelling time-series data, natural language processing, and other applications where long-term dependencies are crucial because they have the ability to selectively retain or forget information over time. In order to model long-term relationships in sequential data, the central component of LSTM is its cell state, which is responsible for enabling this capability. LSTMs are equipped with three gates, which are used to control the flow of information i e,
where:

- it: It represents input gate
- t: It represents forget gate
- ot: It represents output gate
- σ: It represents sigmoid function
- w_χ: It is weight for the respective gate (χ) neurons.
- ht-1: It is the output of the previous LSTM block (at timestamp t-1)
- χt: It is the input at current timestamp
- b_χ: It represents biases for the respective gates (χ)

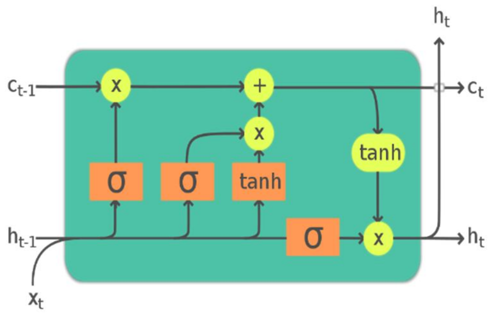

Fig. 2. Working of LSTM

IV. Implementation And Tools

A. Dataset Resources

The dataset for this project was obtained from Yahoo Finance and includes stock data from four companies: JP Morgan Chase & Co., Bank of America Corporation, Wells Fargo & Company, and Morgan Stanley. The Y Finance library in Python was used to access this market data, which provides a wide range of information on stocks for many companies. Specifically, the dataset includes the past year stock data for the aforementioned companies and contains information such as Open, High, Low, Close, Adj Close, Volume, and Company name. The table below provides a detailed explanation of the variables in the dataset.

TABLE I. Dataset Description

Attributes Description of the Dataset	
Attributes	*Description*
Open	This pertains to the initial value at which the stock price commenced for a particular day.
High	This refers to the maximum price attained by the stock during the day.
Low	This refers to the minimum price level that the stock has reached on a given day.
Close	This refers to the closing price of the stock at the end of the trading day.
Adj Close	This variable represents the stock price adjusted for corporate actions like stock splits that impact the stock price. It is a valuable metric for analyzing past returns.
Volume	It represents the total number of securities that were bought and sold during the time interval between the stock market's opening and closing.
Company_ name	This variable indicates the name of the company to which the stock price details pertain.

B. Framework Used

Tensor Flow and Kera's are popular libraries for deep learning that can be used in time series analysis on the stock market. They provide powerful tools for building and training neural networks, which can be used to model complex relationships in time series data. In particular, they can be used for tasks such as prediction, classification, and anomaly detection. For example, LSTM and other RNN models can be built and trained for stock price prediction and other forecasting tasks, and deep autoencoder models can be used for anomaly detection. Transfer learning techniques can also be leveraged to apply pre-trained models for time series analysis on the stock market.

C. Making Data Stationary

In time series analysis for stock market data, ensuring data stationarity is a crucial step. Stationarity implies that the statistical characteristics of the data remain consistent over time. This is important because many time series analysis methods assume that the data is stationary.

There are various methods to transform non-stationary data into stationary data, which includes:

a) *Differencing*: To make the data stationary in time series analysis, one approach is to perform differencing, which involves calculating the difference between the current observation and the previous observation in the time series. This can be done once or multiple times until the resulting data is stationary.

b) *Log transformation*: This can be used to stabilize the variance of the time series data, which can be helpful in making the data stationary.

c) *Seasonal adjustment*: This involves removing seasonal components from the time series data, such as monthly or quarterly seasonal patterns.

d) *Removing trends:* Trends can be removed by fitting a regression model to the time series data and subtracting the fitted values from the data.

D. Shifting the Data

Shifting the data involves Lagging the time series data by a certain number of time periods. This can be useful for several reasons, such as identifying trends, seasonality, or autocorrelation in the data. In the context of stock market analysis, lagging can help identify patterns and trends in the stock prices and trading volumes. For example, you can lag the stock prices by one day to identify if there is any relationship between the current day's stock prices and the previous day's prices.

V. EXPERIMENTAL AND ANALYSIS

A. Auto-Regressive Integrated Moving Average (ARIMA)

To specify the ARIMA model, three parameters are required, namely p, d, and q. The parameter p indicates the number of lagged observations included in the model, the parameter d represents the degree of differencing applied to make the data stationary, and the parameter q represents the size of the moving average window. By using the optimal values of p, d, and q, auto ARIMA generates a fitted model instead of relying on a trial-and-error approach. In our project, we utilized auto ARIMA which gave us the values of p2, d1, and q0. As observed, the value of 0 signifies that we excluded q from the model.

Fig. 3. ARIMA Prediction

TABLE II. ARIMA MODEL RESULT

The results obtained from ARIMA	
Metric	*Value*
Mean Squared Error (MSE)	0.0203
Mean Absolute Error (MAE)	0.0938
Root Mean Squared Error (RMSE)	0.1447

B. Long Short-Term Memory (LSTM)

The LSTM model comprises of two LSTM layers and two dense layers. It takes an input shape of 60*128 and produces a single valued output from the final dense layer. The model has over 117k trainable parameters.

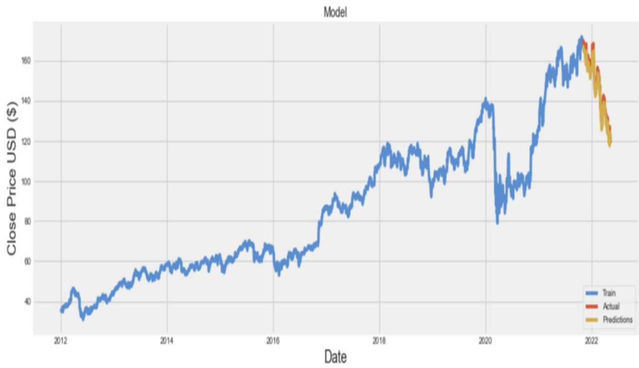

Fig.4. LSTM Prediction

TABLE III LSTM MODEL RESULT

The results obtained from LSTM	
Metric	*Value*
Mean Squared Error (MSE)	8.206
Mean Absolute Error (MAE)	2.206
Root Mean Squared Error (RMSE)	2.874

In comparison to the traditional model ARIMA, we predicted that the neural network models LSTM would produce more accurate results. Surprisingly though The LSTM models had substantial inaccuracy when compared to the conventional ARIMA model, according to the root mean squared error we determined for all the models taking into account the actual closing price and the forecasted closing price.

Whereas the LSTM models had errors of 2.87 and ARIMA model produced a root mean square error of 0.14. This poses the new question of why a classical model would be superior to a neural network model. So, closing price serves as both our input and our prediction for the models' output, which is closing price. We could see from the price prediction graphs that there are differences between the prediction and the actual price when there is a sharp decline. There are a few situations where the actual closing price changes abruptly, either rising or falling. This causes a significant departure from the expected trend line. This results in the enormous variation in the errors between the models.

VI. CONCLUSION AND FUTURE SCOPE

In this Paper, we learned about the many time-series analysis approaches, including both the older statistical techniques and the more recent deep-learning algorithms. We gained knowledge. Some designs preserve long-term memory (LSTM), whilst other architectures solely use short-term elements to forecast (RNN). The exploratory investigation that may be carried out prior to model implementation piqued our interest. In the investigation of stocks, their returns, volumes traded, correlations, and return versus risk, we found a number of data visualizations to be quite helpful.

The deep learning models' implementation satisfied our theoretical understanding, as well. By actually examining the model's layers and trainable parameters, we were able to see why deep learning models, which are built on solid theoretical underpinnings, can predict outcomes with such accuracy.

To compare how well a deep learning model would perform against a traditional model like ARIMA or LSTM, our initial goal was to employ one. In contrast to a usual stock trend line, we discovered that ARIMA was a straight-line forecast. Yet, this performed better than LSTM and estimates of the trend curve, which were quite similar.

The future research will aim to investigate the factors contributing to the superior predictive performance of classical models over neural network models. The study will also seek to identify the variables influencing these predictions and continue to enhance the accuracy of neural network models. This could include incorporating market data or related stock data into the model, there is a great potential for using time series analysis techniques for forecasting future market trends and making investment decisions. Here are some potential future scopes for time series analysis in the stock market:

Incorporating unstructured data: With the increase in social media usage and the availability of news articles, incorporating unstructured data into time series analysis can help to capture market sentiment and provide valuable insights for forecasting.

Advanced machine learning techniques: Deep learning models such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) have shown great promise in time series forecasting. These models can capture complex patterns in the market data and provide more accurate predictions.

High-frequency trading: As technology advances, high-frequency trading (HFT) is becoming more prevalent in the stock market. Time series analysis can be used to identify patterns in high-frequency trading data and develop trading strategies.

Forecasting for emerging markets: Time series analysis can be used to forecast trends in emerging markets, which can be more volatile and unpredictable than developed markets. This can provide valuable insights for investors looking to diversify their portfolios.

Incorporating macroeconomic data: The stock market is influenced by various macroeconomic factors such as inflation, interest rates, and GDP. Incorporating macroeconomic data into time series analysis can help to improve forecasting accuracy.

Thus, the future scope for time series analysis in the stock market is vast and varied, with potential applications in advanced machine learning, high-frequency trading, emerging markets, and macroeconomic forecasting.

ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to everyone who contributed to the successful completion of this research paper on "Time Series Analysis on Stock Market and Forecasting ".

First and foremost, we would want to express our gratitude to our supervisor, Ajay Pal Singh, who has been a huge help to us throughout the study process. His insightful critique and steadfast support have had a significant impact on how our work is progressing.

We would also want to thank Chandigarh University for providing us with the resources and facilities we required for the study.

REFERENCES

[1] Bao, W., Yue, J., & Rao, Y., A deep learning framework for financial time series using stacked autoencoders and long short term memory. PloS one, 12(7), e0180944 (2019).

[2] Liang, Xiaodan, Zhaodi Ge, Liling Sun, Maowei He, and Hanning Chen. (2019) "LSTM with wavelet transform based data preprocessing for stock price prediction." Mathematical Problems in Engineering 2019 .

[3] Naeem, M., Farid, D. M., & Wahab, A. (2021). A hybrid model of grey wolf optimizer and LSTM for stock price prediction. Journal of Ambient Intelligence and Humanized Computing, 12(3), 2381-2393.

[4] Mehtab, S., & Sen, J. (2019). “A robust predictive model for stock price prediction using deep learning and natural language processing.” SSRN 3502624.

[5] Dai, Yeming, Qiong Zhou, Mingming Leng, Xinyu Yang, and Yanxin Wang. (2019) "Improving the Bi-LSTM model with XGBoost and attention mechanism: A combined approach for short-term power load prediction." Applied Soft Computing 130 : 109632.

[6] Li, Y., Liu, H., & Wang, F. Y. (2019), "Stock price and trading volume prediction using a multi-task learning framework with the incorporation of news sentiment". Journal of Intelligent & Fuzzy Systems, volume 36, pages 2967-2980

[7] Ojo, S. O., Owolawi, P. A., Mphahlele, M., & Adisa, J. A. (2019). Stock Market Behaviour Prediction using Stacked LSTM Networks*. International Multidisciplinary Information Technology and Engineering Conference (LIMITED), 2019.

[8] S. Alhazbi, A. B. Said and A. Al-Maadid, "Using Deep Learning to Predict Stock Movements Direction in Emerging Markets: The Case of Qatar Stock Exchange," (2020), IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2020, pp. 440-444, doi: 10.1109/ICIoT48696.2020.9089616.

[9] Zaremba, A., & Świerczyna, D. (2020), "Forecasting Stock Prices with Time Series Analysis and Machine Learning Techniques: A Comparison of Different Approaches", the Journal of Risk and Financial Management, volume 13, issue 10, article 224.

[10] Du, Y., Li, J., & Shang, P. (2021). A comparative study of machine learning and deep learning algorithms for stock price trend prediction. Journal of Computational Science, 50, 101407.

[11] L. Owen and F. Oktariani, "SENN: Stock Ensemble-based Neural Network for Stock Market Prediction using Historical Stock Data and Sentiment Analysis," 2020 International Conference on Data Science and Its Applications (ICoDSA), Bandung, Indonesia, 2020, pp. 1-7, doi: 10.1109/ICoDSA50139.2020.9212982.
[12] Gao, Y., Yan, H., & Jiang, Y. (2020), "Stock Pri-ce Prediction Based on Deep Belief Network Using Multiple Technical Indicators". Risk and Financial Management, volume 13, issue 7, article 151.
[13] Qiao W, Yang Z. Forecast the electricity price of US using a wavelet transform-based hybrid model. Energy. 2020 Feb 15;193:116704.
[14] Tao, Meiyao, Shanshan Gao, Deqian Mao, and Hong Huang. "Knowledge graph and deep learning combined with a stock price prediction network focusing on related stocks and mutation points." Journal of King Saud University-Computer and Information Sciences 34, no. 7 (2022): 4322-4334.