Proceedings of ICFTBA 2024 Workshop: Finance's Role in the Just Transition
DOI: 10.54254/2754-1169/144/2024.GA19085

NASDAQ Forecast Based on ARIMA and ETS Models

Songyu Cai [1, a, *]

[1] *College of Letters, Arts and Sciences, University of Southern California, Los Angeles, US*
a. songyuca@usc.edu
** corresponding author*

Abstract: In the financial markets, stock price prediction has always been a classic, popular, and challenging research topic. The ability to effectively predict stock movements can allow investors to maximize returns and manage risks, which reinforces the efficiency of capital markets. Furthermore, understanding and improving prediction models contributes to the academic study of market behavior and the development of more robust financial models. However, stock prices are subject to various economic, political, and psychological factors, introducing high uncertainty and making accurate forecasting a complex endeavor. In this research, two forecasting methods will be used to forecast the NASDAQ index. Their prediction result will then be compared. The ARIMA model and the ETS model are the two forecasting models used in this study. The past 10 years (2013-2024) of NASDAQ index (open) will be used as the data for the training set and test set. The goal of this study is to compare and evaluate the accuracy of each model and determine the model that is better suited for financial forecasting. The result shows that the RMSE value of ARIMA model fitting is lower than ETS model fitting. The ARIMA model provides a better predicting result in predicting the NASDAQ index. This paper compares ARIMA and ETS models to provide insights for future investment decisions and show a new perspective on stock price forecasting.

Keywords: ARIMA, ETS, NASDAQ, Time series, Prediction

1. Introduction

The NASDAQ Index is one of the most widely tracked and influential stock market indices globally. A broad range of companies listed on the NASDAQ stock exchange, particularly in the technology sector, are being represented by this index. Unlike other stock market indexes that may focus on a subset of industries or larger companies, the NASDAQ Composite includes nearly all securities listed on the NASDAQ, making it a comprehensive measure of market performance. Studying the NASDAQ Composite Index is crucial as it provides insights into the broader market trends, especially in technology and growth-oriented sectors, reflecting changes in market sentiment, economic conditions, and investment patterns.

Previous research on major stock market indices such as NASDAQ index and S&P 500 to predict stock price forecasting has primarily focused on predicting stock prices using approaches that can be split into two groups. The first type is to develop or use a single model. For instance, using a hybrid deep learning based predictive model. It combines Bidirectional Cuda Deep Neural Network LSTM (Long Short-Term Memory) and a one-dimensional Convolutional Neural Network (CNN) [1].

DOI: 10.54254/2754-1169/144/2024.GA19085

Furthermore, some research applies artificial neural networks and build neutral models to predict price behavior [2]; some use stochastic models to describe the dynamics in order to predict short-term stock price behaviors [3]. Lastly, previous research that focus on single model also use novel model for stock price prediction with an integrated multi-stage structure, or a long short-term memory (LSTM) to composite a hybrid bidirectional LSTM and CNN architecture (CNN-BiSLSTM) to predict the closing price of the stock [4-6]. The second type is to use machine learning methods. For example, some research uses extreme learning machine (ELM) [7]; multilayer graph-based learning approach for stock price [8]; study predictive performance of the machine and deep learning methods [9]; use a hybrid model for stock price prediction through learning approach with a sentiment analysis model [10]. Despite the advancements in modeling approaches, stock prediction remains limited in its accuracy due to the fact that financial markets are dynamic and complex. Previous research mostly focuses on a single model. For financial and statistical research, studying and comparing multiple models constructed from the same training dataset is essential and can provide insight for future research. This paper aims to forecast the NASDAQ Composite Index using historical data from 2013 to 2024. The study will apply the ARIMA and the ETS (Error Trend and Seasonality, or exponential smoothing) models, and comparatively analyze their prediction effectiveness. In this paper, the ARIMA and ETS model will be built to forecast the NASDAQ Index (open) time series data. The data from 2013-2023 will be used as the training set, and the 2023-2024 NASDAQ Index will be the test set. Based on the result after fitting the training set to both models, the two models are capable of accurately capturing the overall trend of stock price movements. To further assess their performance, the RMSE (Root Mean Squared Error) of both models was compared in this research, with the ARIMA model demonstrating a smaller RMSE than the ETS model, indicating superior predictive accuracy.

2. Data

The data used in this research spans from 2013 to 2024 with daily prices being collected to improve data continuity and readability. The source of the data being used in this article is from the Federal Reserve Economic Data (FRED). Figure 1 shows the NASDAQ index's trend over this period. The graph highlights a strong uptrend, with noticeable fluctuations. Up to 2020, the index followed a steady growth path, but after 2020, there was a sharp increase, followed by periods of high volatility. The peak levels were observed near the end of the sample in 2024, illustrating significant market shifts during this period. This trend reflects the rapid expansion and subsequent fluctuations in the tech sector, driven by economic conditions and market confidence in innovation-led industries.

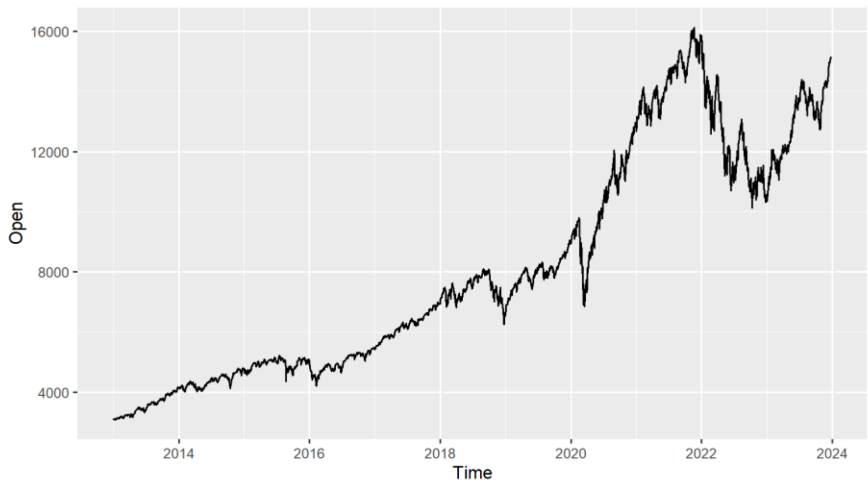

Figure 1: NASDAQ Index (2013-2024)

DOI: 10.54254/2754-1169/144/2024.GA19085

Table 1 shows the basic information of the data used in this paper.

Table 1: Descriptive Statistics of NASDAQ Index

Mean	Standard deviation	Max.	Min.	kurtosis	skewness
8109.9353	3686.0542	16057.44	3091.81	-1.08	0.5218

Table 1 is the basic descriptive statistical characteristics of the data in the selected sample period. According to Table 1, the maximum value of the NASDAQ open prices (2013 January –2024 January) is 16,057.44, while the minimum value is 3,091.81. The mean value for this data set is 8,109.94, and the standard deviation is 3,686.05. This indicates that the data distribution is relatively spread out, reflecting the fluctuations that occurred over the sample period. The kurtosis is -1.08, which is less than 0. This suggests that the distribution is relatively flat compared to a normal distribution exhibiting a lower peak. According to Table 1, the skewness is 0.52. Because it is greater than zero, it shows that the data is slightly skewed to the right, meaning that higher values are more prevalent in the distribution.

3. Methodology

3.1. ARIMA

The ARIMA model is utilized for its ability to capture the underlying patterns in time series data, particularly its capability to model trend and autocorrelation structures effectively. Short term dependencies and patterns in time series data can be effectively captured by prediction using the ARIMA model. It accounts for trends, autoregressive behavior, and moving average of past errors. Through three components, the ARIMA model captures and forecasts dynamic patterns in a dataset. First is the autoregressive component, which models the relationship between the historical value and current value. Second, the differencing component stabilizes the time series by removing trends and transforming nonstationary data into stationary data. Lastly, capturing the relationship between the current values and historical forecast errors (See Equation (1).

$$Y_t = c + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{i=1}^{q} \theta_t \epsilon_{t-i} + \epsilon_t \tag{1}$$

The formula shown above is the formula for ARIMA. For this formula, Yt represents the value of the time series at time t. ϕ1, …, ϕp are the autoregressive parameters that describe the relationship between the current value and its lagged values. The error term ϵt, captures the residual fluctuations not explained by the model. θ1, …, θq in the formula above are the moving average parameters. They capture the effects of past historical errors on the current value. The formula shows that the ARIMA model is a versatile model for forecasting different types of time series data. Furthermore, it is more flexible in capturing autoregressive and moving average components in time series data.

3.2. ETS

The ETS model is designed to predict time series data by decomposing it into three primary components: error, trend, and seasonality. Compared to other forecasting models, the ETS model is relatively simple and flexible, which makes it easy to apply. ETS does not require as many complex parameters unlike the ARIMA model. This reduces the risk of overfitting and makes the model more stable and adaptable. In general, the ETS model is useful for time series with strong seasonal patterns.

DOI: 10.54254/2754-1169/144/2024.GA19085

The ETS model follows a process that involves identifying the components of the time series. It determines whether the series has a trend or seasonality. Then, select the appropriate type of error structure. Lastly, the model is fitted to historical data to estimate its parameter, and further analysis can be conducted (See Equation (2)).

$$yt = (lt - 1 + bt - 1) \times \epsilon t \tag{2}$$

In the formula shown above lt−1 represents the level, and bt−1 is the trend. in this formula represent the multiplicative error term. This equation shows that the forecast is based on the level lt−1. The trend bt−1 is scaled by the multiplicative error term ϵt.

3.3. Comparison

The results from both models will be visualized. Their performance in terms of accuracy is assessed by comparing the RMSE values. This will determine which model offers better prediction for the NASDAQ Index. The analysis will provide insight into the applicability of time series data forecasting methods in predicting complex financial data.

4. Results

4.1. ARIMA

For more reliable forecasting, a stationary time series is needed for ARIMA modeling. This will ensure consistent statistical properties over time. The null hypothesis is set as that the time series is nonstationary (or it has a unit root) in the context of the ADF (Augmented Dickey Fuller) test. Strong statistical evidence against the null hypothesis that suggests that the time series is stationary will be indicated if the p-value in the ADF test is less than 1 percent, 5 percent, or 10 percent. For the data used in this paper, the ADF test results show a p-value that is approximately 0 (< 2e-16). It is less than the critical values at 1 percent, 5 percent, and 10 percent significance levels. This shows that there is statistical evidence to reject the null hypothesis. Therefore, the time series data used in this paper is stationary.

The ACF (Autocorrelation Function) graph shows the direct relationship of a time series with its own lagged values. For the first order difference log-transformed data, the ACF at lag 0 is 1. It represents the correlation of the series with itself by definition. The autocorrelation values quickly drop close to 0 after lagging 0, t. This shows that there are no strong relationships between the series values. Furthermore, its past values are beyond the immediate lags. There is a rapid decline in autocorrelation. This suggests that the differenced series is a stationary white noise process, which is a desirable property for ARIMA modeling (See Figure 2).

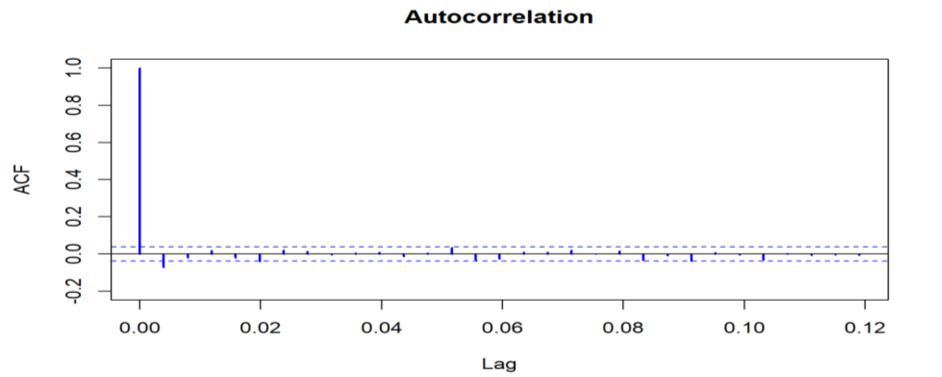

Figure 2: Autocorrelation

DOI: 10.54254/2754-1169/144/2024.GA19085

Figure 3 is a partial auto correlation plot (PACF). It shows a strong positive spike at lag one. This shows the correlation of the current value and its immediate past value. Additionally, all other lags fall within the confidence intervals. This represents that there are no significant correlations beyond lag one. Lastly, this pattern suggests that for the data used in this paper, an ARIMA (1,1,0) model is appropriate, as the series has a first-order autoregressive component after differencing.

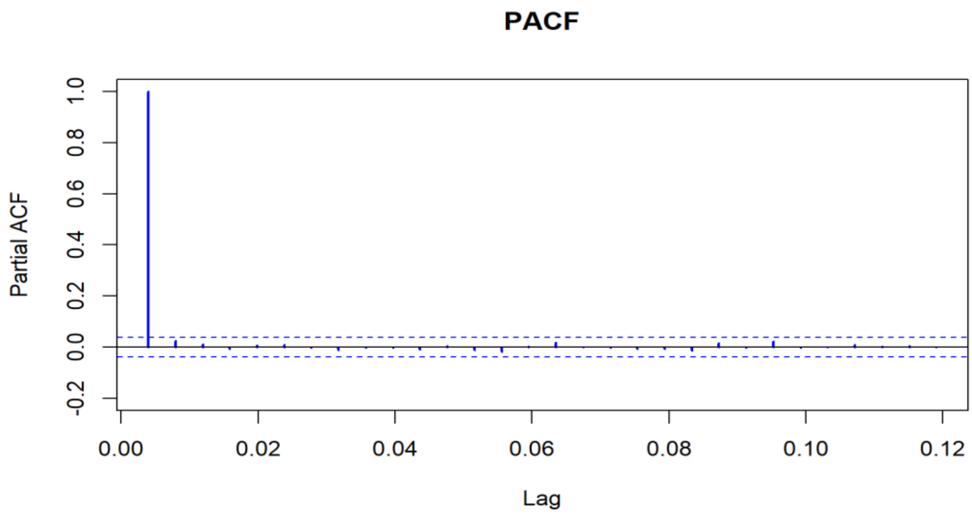

Figure 3: Partial Autocorrelation

The model used in this analysis is an ARIMA (0,1,1) with drift. Composition of the drift term indicates a linear trend in the differenced series. The coefficients for this model are estimated at -0.0760, the standard error is 0.0202, and the drift term at 0.0005 with a standard error of 0.0002. The model test is carried out, and the p values is less than the set probability threshold of 0.05. This shows that it is statistically significant, and the null hypothesis is rejected, which means that the model diagnosis passes.

The model fit is assessed using AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values. The AIC for this model is -14790.38, and the BIC is -14772.88. The large negative value could result from the very high log-likelihood value, which happens with large dataset or model that fits the data extremely well. The auto. Arima () function selects the model with the smallest BIC and AIC, indicating that this ARIMA (0,1,1) model with drift was chosen because it minimizes these criteria, suggesting a good model fit.

The model predicts and visualizes the NASDAQ index(open) from 2023 to 2024. The training set error measures provide an interpretation of the predictive accuracy of the model. The Mean Error, ME is approximately 0 (3.430137e-06), suggesting that the model's forecasts are unbiased on average. The RMSE (Root Mean Squared Error) is a crucial measurement that can be used to show model accuracy. For this data, it is calculated as 2009.58(RMSE of the forecast value, real value). This value represents the standard deviation of the prediction errors, indicating the average magnitude of the errors in the forecasts.

As shown in Figure 4, the forecast suggests a gradual upward trend in the NASDAQ index, with a confidence interval widening as time progresses. This indicates an increasing uncertainty in the predictions for when moving further into the future. The model captures some short-term fluctuations but overall shows a recovery from recent market volatility.

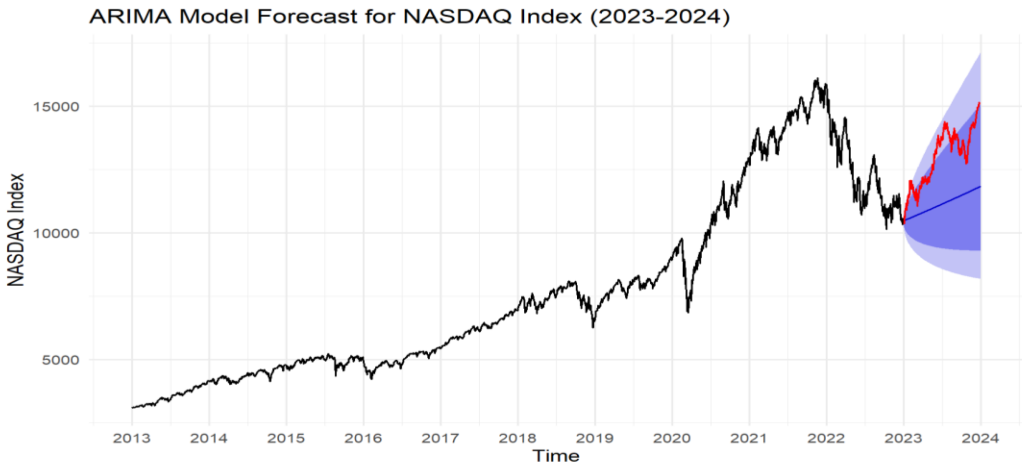

Figure 4: ARIMA Prediction Result

4.2. ETS

The ETS model used in this paper is ETS (M, A, N). It indicates a multiplicative error, additive trend, and no seasonality. The multiplicative error, M, shows that the variability in the forecast increases with the series level proportionally. The additive trend component, A, indicates that the trend in the data is linear over time. Lastly, the absence of a seasonal component, N, shows that the model does not capture any regular periodic fluctuations in the NASDAQ Index data. The smoothing parameters show α = 0.914, and β = 1e-04, suggesting that the model provides significant weight to new observations, when estimating the series level, and the model considers the trend component to be comparably constant or stable over time. As shown in Figure 5, the model predicts and visualizes the NASDAQ index(open) from 2023 to 2024. The ETS model captures some of the short-term fluctuations and shows a predicted upward trend. Compared to the real NASDAQ Index data, the ETS model seems to provide reasonable forecasts, but it struggles with the inherent volatility of the stock prices and the financial market. This shows that while ETS can predict general trends, its accuracy decreases as the forecast horizon increases, and it does not fully capture the more volatile swings in the index.

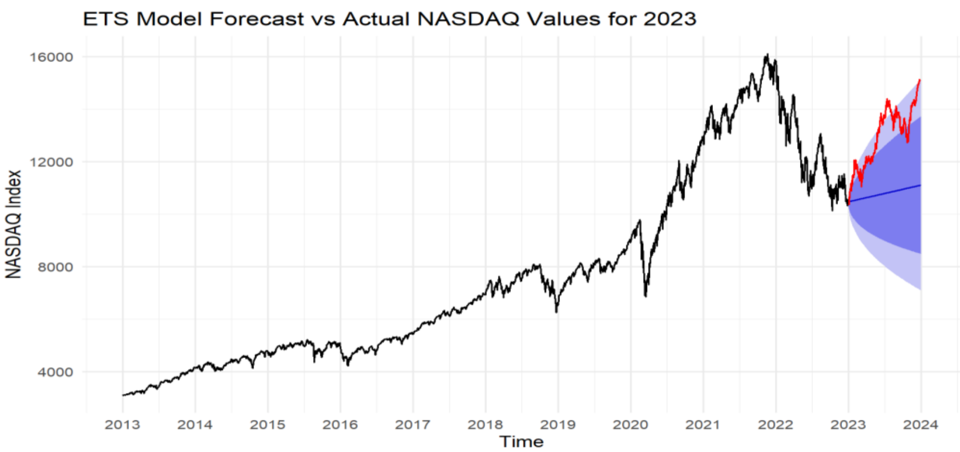

Figure 5: ETS Prediction Result

4.3. Comparison

Comparing the predictions of the two models: Figure 6 and Table 2, the conclusion can be made that ARIMA model is superior in accuracy than the ETS model in predicting the performance of the NASDAQ Index. This is further supported by the RMSE index of the two models. For the result, the RMSE for the ETS model is approximately 2389.47. It is higher than the RMSE of the ARIMA model, which is approximately 2009.58. According to the RMSE, the ARIMA model has a slightly better predictive accuracy for the NASDAQ index data compared to the ETS model.

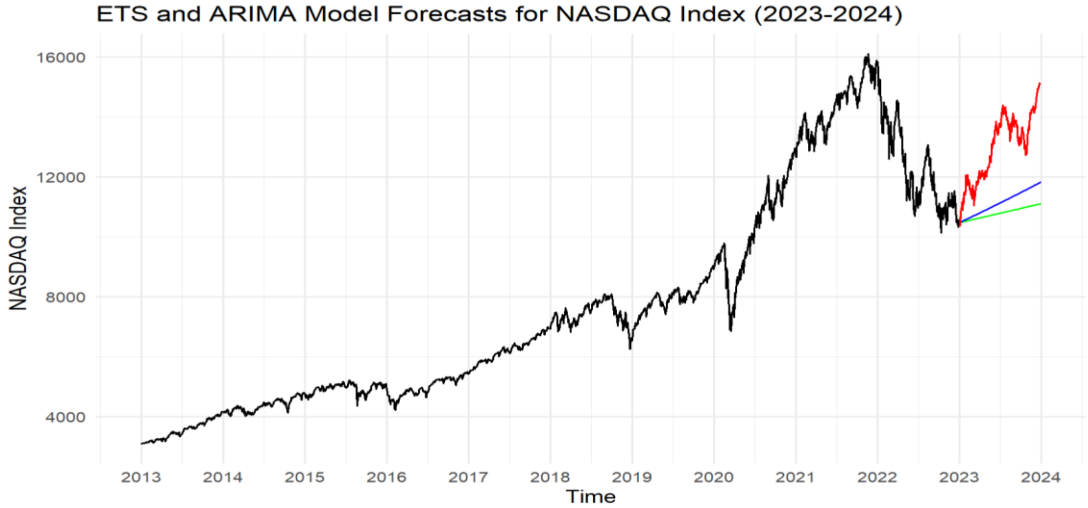

Figure 6: ETS, ARIMA Prediction Results Comparison

ETS Prediction = green, ARIMA Prediction = blue, Actual 2023-2024 NASDAQ Index = red

Table 2: RMSE Comparison Results

Model	RMSE results
ARIMA	2009.5817
ETS	2389.4720

5. Conclusion

In the financial markets, stock prices forecasting had been a fundamental, yet difficult challenge. Investors will have a better ability to optimize returns and mitigate risks if forecasting stock price movements can be accurately made. Furthermore, it would enhance the overall efficiency of capital markets. This research uses ARIMA and ETS models to forecast the NASDAQ Index from 2023 to 2024. The prediction effect of the two models is compared and analyzed. The RMSE value of the two models is used as the standard of prediction. According to the result of the comparison, the ARIMA model provides better prediction results relative to ETS model. In conclusion, ARIMA provides more precise predictions by effectively capturing the patterns and dynamics of time series data. This makes it a useful tool for investors aiming to optimize returns and manage risk. This paper could provide insight into future investment strategies and contribute to the development of more accurate models for market prediction. Furthermore, it expands the horizon of asset price forecasting and provides a new way of thinking for asset price forecasting.

In future research, researchers could understand market dynamics more comprehensively through exploring the NASDAQ Index or other major indices from two dimensions. First, researchers could

DOI: 10.54254/2754-1169/144/2024.GA19085

broaden the time span of the research or use higher frequency dataset. This would allow forecast indices from a more macro perspective. Second, to experiment with multiple different forecasting models. This would discover the model that performs best in predicting stock market indices. Through these approaches, more accurate asset price forecasts can be provided from researchers to investors.

References

[1] *Kanwal, A., Lau, M. F., Ng, S. P. H., Sim, K. Y., & Chandrasekaran, S. (2022). BiCuDNNLSTM-1dCNN — A hybrid deep learning-based predictive model for stock price prediction. Expert Systems with Applications, 202, 117123*

[2] *de Oliveira, F. A., Nobre, C. N., & Zárate, L. E. (2013). Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4, Petrobras, Brazil. Expert Systems with Applications, 40(18), 7596–7606.*

[3] *An, Y., & Chan, N. H. (2017). Short-Term Stock Price Prediction Based on Limit Order Book Dynamics. Journal of Forecasting, 36(5), 541–556.*

[4] *Sedighi, M., & Roodposhti, F. R. (2022). Designing a Novel Model for Stock Price Prediction Using an Integrated Multi-Stage Structure: The Case of the Bombay Stock Exchange. Australasian Accounting, Business & Finance Journal, 16(6), 70–85.*

[5] *Chaudhary, A., Gupta, A., Pahariya, D., & Singh, S. K. (2023). Stock Price Prediction of Tesla & Apple using LSTM. ITM Web of Conferences, 56, 2006.*

[6] *Wang, H., Wang, J., Cao, L., Li, Y., Sun, Q., & Wang, J. (2021). A Stock Closing Price Prediction Model Based on CNN-BiSLSTM. Complexity, 1.*

[7] *Boru İpek, A. (2023). Stock price prediction using improved extreme learning machine methods during the Covid-19 pandemic and selection of appropriate prediction method. Kybernetes, 52(10), 4081–4109.*

[8] *Alshara, M. A. (2022). Multilayer Graph-Based Deep Learning Approach for Stock Price Prediction. Security and Communication Networks, 2022, 1–8.*

[9] *Lachaab, M., & Omri, A. (2023). Machine and deep learning-based stock price prediction during the COVID-19 pandemic: the case of CAC 40 index. EuroMed Journal of Business.*

[10] *Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. Expert Systems with Applications, 178, 115019.*