Journal of Business and Management Studies
ISSN: 2709-0876
DOI: 10.32996/jbms
Journal Homepage: www.al-kindipublisher.com/index.php/jbms

JBMS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Dominance of External Features in Stock Price Prediction in a Predictable Macroeconomic Environment

Md Tuhin Mia[1], Rejon Kumar Ray[2], Bishnu Padh Ghosh[3], Md Salim Chowdhury[4], Md Al-Imran[5], Malay Sarkar[6], Radha Das[7], Nilufer Sultana[8] ✉ and Saad Abrar Nahian[9] and Aisharyja Roy Puja[10]

[13]School of Business, International American University, Los Angeles, California, USA
[6]Department of Management Science and Quantitative Methods, Gannon University, USA
[2]Department of Business Analytics, Gannon University, USA
[45]College of Graduate and Professional Studies, Trine University, USA
[7]Researcher, Dhaka, Bangladesh
[8]Researcher, Centennial, CO, USA
[9]Researcher, University of Hertfordshire, Hertfordshire, UK
[10]Management Science and Quantitative Methods, Gannon University, USA

Corresponding Author: Nilufer Sultana, **E-mail**: nilufersultana@gmail.com

| ABSTRACT

Understanding the factors affecting future stock prices has been of prime importance across the globe, as accurate stock price prediction is directly related to financial gains. Its interest has been reflected by a large and growing literature trying to investigate stock price prediction with an effort to gain higher prediction accuracy. Recent literature has identified relevant external features, such as current and anticipated future macroeconomic environment-related information, and has incorporated such external features along with historical data on stock prices into the prediction models to gain improved accuracy. However, the current literature fails to quantify the relative importance of those external features for a better understanding of their relevancy. In this article, we bridge this gap and quantify the relative importance of those external features in stock price prediction by combining macroeconomic data with historical stock price data and by utilizing dominance analysis. Our results demonstrate that external features are highly dominant in the prediction of future stock prices.

| KEYWORDS

Dominance Analysis, Relative Feature Importance, Stock Price Prediction

| ARTICLE INFORMATION

ACCEPTED: 01 December 2023 **PUBLISHED:** 21 December 2023 **DOI:** 10.32996/jbms.2023.5.6.10

1. Introduction

In this paper, we investigate the contribution of the current and anticipated future macroeconomic environment in the stock price prediction problem using dominance analysis. Specifically, we quantify the amount of predicted variance the current and anticipated future macro policy is responsible for in stock price predictions in a predictable macroeconomic environment, thereby shedding light on the significance of the current and future external environment in stock price prediction. Our findings unveil new insights into the driving forces of future stock prices, thereby assisting investors in optimizing their investment strategies.

Stock price prediction has gained significant attention as it directly affects financial gains. Stock price prediction is a delicate problem as the stock market is highly volatile and sensitive to external factors. This sensitivity of the stock market to the external macroeconomic condition has been well-documented in the literature [Diebold and Yilmaz, 2008]. This delicate nature of the stock market invokes special attention, and therefore, a growing number of research efforts have been carried out to investigate stock prices. Recent literature has focused on the inclusion of external factors that can explain the future movement of stock prices. Haque, Amin, Miah, Cao and Ahmed, 2023 cleverly demonstrated that anticipated macro policy, along with other macroeconomic features, can affect future stock prices, and the inclusion of such features in the prediction model significantly improved the accuracy of the prediction performance. They demonstrate such relevancy by comparing prediction performances with and without such anticipated macro policy along with other macro variables. While there is very little evidence of performance enhancement of stock price prediction due to the inclusion of external factors, we are not aware of any study quantifying the relative strengths of such features in stock price prediction problems. We employ dominance analysis and quantify the relative importance of such external features, thereby further confirming their relevancy. We combine historical stock price data with macroeconomic variables and perform dominance analysis to quantify the relative importance of each feature and then rank each feature in order of magnitude of relative importance. We further confirm our results by investigating important features identified by Shapely values. Our results suggest that such identified external features are highly dominant, thereby implying that the inclusion of such features will likely improve prediction accuracy.

2. Literature Review

Past investigations in this domain have predominantly relied on conventional statistical approaches. Seber and Lee, 2013 employed linear regression to predict future stock prices]. Other studies have utilized various autoregressive linear time series models to predict future stock prices. Zhang (2003) relied on the Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) for stock price predictions. Other approaches considered in the stock price prediction literature include Random Walk Theory (RWT) [Reichek and Devereux, 1982] and Moving Average Convergence/Divergence (MACD) [Chong and Ng, 2008] for forecasting stock prices. With the development of advanced algorithms in the area of machine learning and artificial intelligence, contemporary research has shifted its focus towards leveraging such advanced algorithms, given their superior performance in stock price prediction. Researchers have explored tree-based models like Random Forest (RF) [Liaw & Wiener, 2002], as well as neural network-based approaches such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) [Li et al., 2017, Oyeyemi et al., 2007]. ANN, with its capability to extract latent features through self-learning, stands out as a suitable choice for stock price prediction. Its strength as a potent approximator enables it to discern complex input-output relationships within extensive datasets, making ANN a viable option for forecasting an organization's stock prices.

Mizuno et al., 1998 utilized ANN for technical analysis on the TOPIX dataset, implementing it in a system for predicting buying and selling timing. Some studies have suggested the use of Random Forest (RF) for forecasting, an ensemble technique excelling in regression and classification tasks by constructing multiple decision trees during training, providing a mean regression of individual decision trees [Kumar and Thenmozhi, 1996]. Roman et al. applied RNN models to stock market data from five countries—Canada, Hong Kong, Japan, the UK, and the USA—to predict trends in stock returns [Roman and Akhtar, 1996]. Rout et al., 2017 utilized a simplified RNN model for predicting the stock market, evaluating its performance on datasets from the Bombay Stock Exchange and the S&P 500 index. Hamzaçebi et al., 2009 delved into multi-periodic stock market forecasting using the ANN model. Selvin et al., 2017 compared various deep-learning techniques in their study on stock price predictions for NSE-listed companies. Yunus et al., 2014 employed ANN on NASDAQ data to forecast stock closing prices. Mei et al., 2014 effectively applied RF to predict real-time prices in the New York electricity market. Herrera et al., 2010 leveraged RF as a predictive model for forecasting hourly urban water demand. In a more recent development, Khan et al., 2023 utilized reinforcement learning algorithms to predict stock prices.

While a considerable body of literature exists on stock price predictions, it largely overlooks the potential impact of anticipated future macroeconomic policy changes and economic conditions. Instead, the focus has predominantly been on utilizing historical stock price data to develop forecasting models. Currently, existing literature mostly relies on time-series data concerning stock prices, along with other pertinent variables, to construct forecasting models. Nevertheless, as argued by Haque, Amin, Miah, Cao and Ahmed, 2023, it is essential to recognize that broader economic conditions can exert a significant impact on investments. Additionally, certain macroeconomic policy variables, such as interest rates, can be reasonably anticipated in advance, as central banks regularly engage in public discussions regarding their future actions and strategies. Given that interest rates can directly influence investment costs and, consequently, returns, investors have a compelling incentive to form expectations about future interest rate changes based on the central bank's public discussions. Subsequently, they adjust their current investments in accordance with their anticipated future interest rates. Authors in that study demonstrated that prediction performances can be significantly enhanced by including anticipated future macro policy along with other relevant variables related to macroeconomic environments. In another companion research, Alamsyah and Zahir, 2018 explored the use of macroeconomic variables such as inflation rates, interest rates, and exchange rates to predict the IDX Composite Index, which gauges the stock price performance of all listed companies on the Indonesia Stock Exchange. There are a few other studies that investigate the relevance of the macroeconomic environment in retail demand prediction problems and find evidence for the existence of relevancy [Haque (2023); Haque, Amin, Miah, Cao, Sayed and Ahmed, 2023]. Haque (2020) presented evidence supporting intertemporal behavioral adjustments in response to expected future fiscal policy changes. While there has been very little evidence of the relevance of external factors, such as macroeconomic conditions, anticipated macro policy, etc. [Haque, Amin, Miah, Cao and Ahmed, 2023]. To the best of our knowledge, no prior research has attempted to quantify the relative importance of such features in stock price predictions. We contribute to the literature by examining the magnitude of the relative importance of external features in stock price predictions using dominance analysis. Findings from our study confirm the importance of the inclusion of external factors in the prediction of stock prices in a highly volatile market.

3. Methodology

For this study, we compile the dataset by amalgamating information from diverse sources. Historical daily stock price data for S&P500 ticker symbols is gathered from the publicly accessible Yahoo Finance. We specifically chose ten ticker symbols characterized by the highest Partial Autocorrelation (PAC) values, indicating robust serial autocorrelation crucial for capturing temporal dependencies. 7 days lead (ahead) of the daily Close field has been used as the target variable for this analysis. Data on interest rates are sourced from Fred's data repository. Additionally, unemployment rates and Consumer Price Index (CPI) data are acquired from the World Bank's World Development Indicators (WDI) database. These distinct datasets are then consolidated into a unified dataset, with all the information appended together.

We use standard feature engineering techniques to generate the results. First, we generate dummies for each month of the year and day of the week. We use all the current variables, along with fourteen-day lead (future) interest rates as independent variables and seven-step lead (future) stock price (Close field) as the target variable for the analysis. Lead interest rates serve as a proxy for anticipated future interest rates fourteen days ahead of time. We also include 7 and 14-day rolling averages and rolling standard deviations of stock prices. I restrict the analysis to the years 2017-2019 to avoid capturing any trends that may not exist in recent years. All variables included in the model are presented in Table 1.

Table 1: List of Independent Variables

Independent Variables	Descriptions
volume	Number of stocks traded
cpi	Consumer Price Index
unemp	Unemployment rates
int_rate	Current interest rates
lead_int_rate	Anticipated future interest rates 14 days ahead
stock_price	Current stock price
rolling_mean_t7	7 days rolling average of stock price
rolling_mean_t14	14 days rolling average of stock price
rolling_std_t7	7 days rolling standard deviation of stock price
rolling_std_t14	14 days rolling standard deviation of stock price
Indicator variables	Indicators for months of a year, days of a week, and ticker symbols

Dominant features and their relative importance have been determined by using Dominance Analysis. Dominance analysis (DA) is a method that compares the relative importance of predictors in multiple regression. DA determines the dominance of one predictor over another by comparing their additional R^2 contributions across all subset models.

4. Results and Discussions

Table 2 shows the rankings resulting from dominance analysis. The current stock price, lead interest rates, current interest rates, and 7 days rolling average of stock price are found to be the most influential predictors, accounting for 37.01%, 23.19%, 18.26%, and 12.03% of the predicted variances, respectively. Other features, such as volume, cpi, unemployment rates, 14 days rolling average, 7 days rolling standard deviation, 14 days rolling standard deviation, and several indicator variables, are responsible for relatively smaller predicted variances.

Table 2: Dominance importance of predictors

Feature	Standardized Dominance Statistic	Rank
volume	0.0204	5
cpi	0.0196	6
unemp	0.0115	7
int_rate	0.1826	3
lead_int_rate	0.2319	2
stock_price	0.3701	1
rolling_mean_t7	0.1203	4
rolling_mean_t14	0.0112	8
rolling_std_t7	0.0109	9
rolling_std_t14	0.0103	10

Relative importance on several indicators is suppressed for brevity.

To ascertain the relationship between stock prices and both current and future interest rates, along with other macroeconomic variables, a linear regression is employed, estimating all features' impact on stock prices. The estimates from this regression are presented in Table 3. For brevity, coefficients on several variables are omitted. In the table, coefficients on macroeconomic variables and expected future interest rates are displayed, along with their corresponding p-values. Significance is observed at a 10% level for coefficients on current stock price, 7-day rolling average of stock prices, CPI, unemployment rates, current interest rates, and anticipated future interest rates. The coefficient on volume is -0.0031, indicating a statistically significant association wherein higher sales volume correlates with lower stock prices. With a coefficient of 1.113 at a 10% significance level, the CPI is positively linked to higher stock prices, suggesting that an increased inflation rate is associated with elevated stock prices. Similarly, higher unemployment rates exhibit a statistically significant negative association with stock prices. The negative coefficient on current interest rates is statistically significant at a 5% level, aligning with the expectation that rising interest rates make investments less profitable and less attractive, thereby contributing to a decline in stock prices. Anticipated interest rates have positive coefficients, statistically significant at a 10% level, consistent with the notion that investors foreseeing future increases in investment costs, are incentivized to invest before an actual interest rate hike occurs.

Table 3: Estimates from multiple regression on (future) stock prices

Feature	Coefficient	p-value
volume	-0.0031	0.04
cpi	1.113	0.07
unemp	-2.379	0.09
int_rate	-3.135	0.05
lead_int_rate	0.127	0.08
stock_price	0.092	0.01
rolling_mean_t7	0.032	0.08
rolling_mean_t14	0.432	0.23
rolling_std_t7	0.032	0.18
rolling_std_t14	0.117	0.26

Relative importance on several indicators is suppressed for brevity.

The last step of our verification includes an examination of the important features as identified by Shapely (SHAP) values obtained from the fitted multiple linear regression model and comparing them with the dominant features identified by dominance analysis. Shapely values utilize the Collaborative Game theory approach to provide desirable properties and are widely used in literature for explaining computational intelligence models [Chen et al., 2021]. I explain the importance of the proposed external features compared to other features in the model by using Shapely values. The first five important features identified by Shapely values are presented in Figure 1. Findings are similar to those found in the dominant analysis. Like the findings from dominance analysis, the first three important features are current stock prices, current interest rates, and future interest rates. These results further confirm that the current and anticipated future macroeconomic environment is very relevant in stock price prediction.

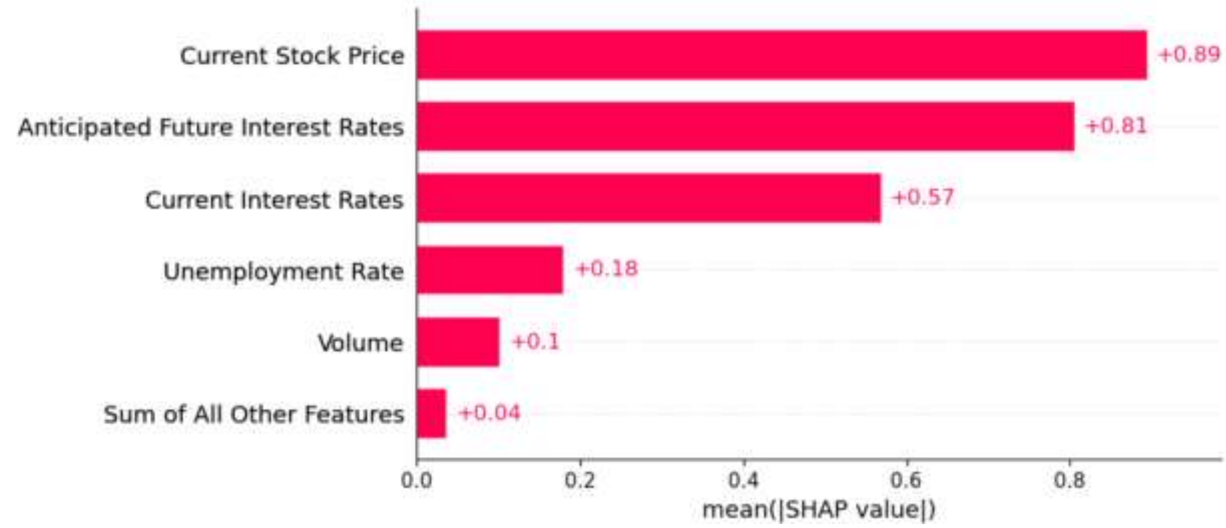

Figure 1: Shapely Value (Feature Importance)

5. Conclusion

In this study, we investigate the dominance of features in stock price prediction. By amalgamating stock market data and macroeconomic data, we find the magnitude of the relative importance of features included in the prediction model. We further confirm these findings by using Shapely values and comparing the important features obtained using Shapely values with those identified by dominance analysis. Our empirical findings confirm that investors are forward-looking, form an expectation about the future macroeconomic environment, and make their current investment decisions considering both the current and future macroeconomic environment.

These findings hold significant potential for both academic research and practical industry applications. This proposed technique has the potential to disrupt the investment market, as the results directly correlate with enhanced return on investment. Investors can employ our suggested method to make more informed and advantageous investment choices. Our findings are based on the US stock market and are limited to short-term predictions. Future research can investigate and confirm these results using other investment markets and varying prediction periods.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Alamsyah, A., & Zahir, A. N. (2018, May). Artificial Neural Network for predicting Indonesia stock exchange composite using macroeconomic variables. In 2018 6th International Conference on Information and Communication Technology (ICoICT) (pp. 44-48). IEEE.

[2] Chen J., Koju W., Xu S., and Liu Z. (2021). Sales Forecasting Using Deep Neural Network and SHAP Techniques. 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 135–138

[3] Chong, T. T. L., & Ng, W. K. (2008). Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30. Applied Economics Letters, 15(14), 1111-1114.

[4] Diebold, F. X., & Yilmaz, K. (2008). *Macroeconomic volatility and stock market volatility worldwide* (No. w14269). National Bureau of Economic Research.

[5] Herrera, M, Lu´ıs T, Joaqu´ın I, and Rafael P. (2010). Predictive models for forecasting hourly urban water demand. *Journal of Hydrology* 387 (1-2): 141-150.

[6] Hamzaçebi, C., Akay, D., & Kutay, F. (2009). Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. Expert systems with applications, 36(2), 3839-3844.

[7] Haque, M. S. (2020). Three Essays in Public Finance.

[8] Haque, M. S. (2023). Retail Demand Forecasting Using Neural Networks and Macroeconomic Variables. *Journal of Mathematics and Statistics Studies, 4*(3), 01-06.

[9] Haque, M. S., Amin, M. S., Miah, J., Cao, D. M., & Ahmed, A. H. (2023). Boosting Stock Price Prediction with Anticipated Macro Policy Changes. *Journal of Mathematics and Statistics Studies*, *4*(3), 29-34.

[10] Haque, M. S., Amin, M. S., Miah, J., Cao, D. M., Sayed, M. A., & Ahmed, S. (2023). Retail Demand Forecasting: A Comparative Study for Multivariate Time Series. *Journal of Mathematics and Statistics Studies*, *4*(4), 40-46.

[11] Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers & operations research, 32(10), 2513-2522.

[12] Hur, J., Raj, M., & Riyanto, Y. E. (2006). Finance and trade: A cross-country empirical analysis on the impact of financial development and asset tangibility on international trade. World Development, 34(10), 1728-1741.
[13] Islam, M. M., Nipun, S. A. A., Islam, M., Rahat, M. A. R., Miah, J., Kayyum, S., ... & Al Faisal, F. (2020). An Empirical Study to Predict Myocardial Infarction Using K-Means and Hierarchical Clustering. In Machine Learning, Image Processing, Network Security and Data Sciences: Second International Conference, MIND 2020, Silchar, India, July 30-31, 2020, Proceedings, Part II 2 (pp. 120-130). Springer Singapore.
[14] Khan, R. H., Miah, J., Rahman, M. M., Hasan, M. M., & Mamun, M. (2023, June). A study of forecasting stocks price by using deep Reinforcement Learning. In 2023 IEEE World AI IoT Congress (AIIoT) (0250-0255). IEEE.
[15] Kumar, M., & Thenmozhi, M. (2006, January). Forecasting stock index movement: A comparison of support vector machines and random forest. In Indian institute of capital markets 9th capital markets conference paper.
[16] Khan, R. H., Miah, J., Arafat, S. M. Y., Syeed, M. M. M., & Ca, D. M. (2023). Improving Traffic Density Forecasting in Intelligent Transportation Systems Using Gated Graph Neural Networks. arXiv preprint, arXiv:2310.17729.
[17] Khan R. H., Miah J., Abed N S. A. and Islam M., (2023). A Comparative Study of Machine Learning classifiers to analyze the Precision of Myocardial Infarction prediction, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, 0949-0954, Doi: 10.1109/CCWC57344.2023.10099059.
[18] Kayyum S. et al., (2020). Data Analysis on Myocardial Infarction with the help of Machine Learning Algorithms considering Distinctive or Non-Distinctive Features, 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, 1-7, doi: 10.1109/ICCCI48352.2020.9104104.
[19] Khan, R. H., Miah, J., Rahman, M. M., & Tayaba, M. (2023, March). A Comparative Study of Machine Learning Algorithms for Detecting Breast Cancer. In *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 647-652). IEEE.
[20] Li, L., Wu, Y., Ou, Y., Li, Q., Zhou, Y., & Chen, D. (2017, October). Research on machine learning algorithms and feature extraction for time series. In 2017 IEEE 28th annual international symposium on personal, indoor, and mobile radio communications (PIMRC) (pp. 1-5). IEEE.
[21] Liaw, A., & Wiener, M. (2002). Classification and regression by Random Forest. R news, 2(3), 18-22.
[22] Masoud, N. M. (2013). The impact of stock market performance upon economic growth. *International Journal of Economics and Financial Issues,* 3(4), 788-798.
[23] Mei, J, Dawei H, Ronald H, Thomas H, and Guannan Q. (2014). A random forest method for real-time price forecasting in New York electricity market. IEEE PES General Meeting Conference & Exposition: 1-5.
[24] Mizuno, H, Michitaka K, Hiroshi Y, and Norihisa K. (1998) Application of neural network to technical analysis of stock market prediction. Studies in Informatic and control 7 (3): 111-120.
[25] Murkute, A., & Sarode, T. (2015). Forecasting market price of stock using artificial neural network. *International Journal of Computer Applications,* 124(12), 11-15.
[26] Miah, J., Cao, D. M., Sayed, M. A., & Haque, M. S. (2023). Generative AI Model for Artistic Style Transfer Using Convolutional Neural Networks. arXiv preprint arXiv:2310.18237
[27] Miah, J., Haque, M. S., Cao, D. M., & Sayed, M. A. (2023). Enhancing Traffic Density Detection and Synthesis through Topological Attributes and Generative Methods. *Journal of Computer Science and Technology Studies*, *5*(4), 69-77.
[28] Miah, J., Ca, D. M., Sayed, M. A., Lipu, E. R., Mahmud, F., & Arafat, S. M. Y. (2023). Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction. arXiv preprint, *2311.00517*
[29] Oyeyemi, E. O., McKinnell, L. A., & Poole, A. W. (2007). Neural network-based prediction techniques for global modeling of M (3000) F2 ionospheric parameter. Advances in Space Research, 39(5), 643-650.
[30] Reichek, N., & Devereux, R. B. (1982). Reliable estimation of peak left ventricular systolic pressure by M-mode echographic-determined end-diastolic relative wall thickness: identification of severe valvular aortic stenosis in adult patients. *American heart journal,* 103(2), 202-209.
[31] Roman, J, and Akhtar J. (1996). Backpropagation and recurrent neural networks in financial analysis of multiple stock market return. Proceedings of HICSS-29: 29th Hawaii International Conference on System Sciences 2: 454-460.
[32] Rout, A Kumar, P. K. Dash, R D, and Ranjeeta B. (2017). Forecasting financial time series using a low complexity recurrent neural network and evolutionary learning approach. *Journal of King Saud University-Computer and Information Sciences 29* (4): 536-552.
[33] Seber, G. A., & Lee, A. J. (2003). Linear regression analysis (330). John Wiley & Sons.
[34] Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017, September). Stock price prediction using LSTM, RNN and CNN-sliding window model. In 2017 International Conference on Advances in Computing, Communications and Informatics (icacci) (1643-1647). IEEE.
[35] Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural Processing Letters, 9, 293-300.
[36] Yetis, Y, Halid K, and Mo J. (2014) Stock market prediction by using artificial neural network. in 2014 World Automation Congress (WAC): 718-722.
[37] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.