JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Forecasting Stock Closing Prices through Artificial Intelligence Approaches

[1]Vikas Burri, [2]Lalasa Mukku
[1]University of Colorado, Fort Collins, USA
[2]CHRIST(Deemed to be University),Bangalore, India

Abstract : Accurately forecasting stock market returns presents a significant challenge due to the unpredictable and complex nature of financial markets. The advancement of artificial intelligence and improved computational capabilities has led to more effective predictive methods for stock prices. This study applies Artificial Neural Network and Random Forest techniques to predict the closing price for five companies operating in different sectors using financial data such as Open, High, Low, and Close prices. New variables derived from this data are used as inputs for the models. Evaluation of the models based on standard strategic indicators like RMSE and MAPE indicates their efficiency in predicting stock closing prices due to their low values.

***IndexTerms* – Machine learning, deep learning, artificial intelligence, stock market, stock trading**

I. Introduction

The stock market is known for its dynamic, unpredictable, and non-linear nature. Forecasting stock prices is a complex endeavor as it relies on multiple factors such as political conditions, the global economy, company financial reports and performance. Therefore, analyzing trends over recent years to predict stock values in advance could be valuable for maximizing profits while minimizing losses in stock market transactions [1],[2]. Two primary methods have traditionally been suggested for forecasting an organization's stock price. The technical analysis approach involves utilizing historical stock prices such as closing and opening prices, trading volume, and adjacent close values to predict future stock prices. Qualitative analysis, on the other hand, is conducted based on external factors like company profile, market conditions, political and economic influences, textual data from financial news articles, social media posts, and even blogs by economic analysts [3]. Nowadays, modern sophisticated methods involving either technical or fundamental analysis are applied to forecast stock prices. In the domain of stock market analysis, the volume of data is substantial and non-linear in nature. To effectively handle this diverse dataset, an efficient model capable of recognizing concealed patterns and intricate relationships within this extensive data set is required. Machine learning techniques have demonstrated a 60-86 percent enhancement in efficiencies when compared to previous methods in this field[4].

The majority of prior research endeavors within this domain have predominantly relied upon conventional methodologies, notably linear regression algorithms[5]. Previous studies have often employed a variety of established methodologies, including Random Walk Theory (RWT) [6], Moving Average Convergence/Divergence (MACD) [7], and have further incorporated linear models such as Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) [8] to forecast stock prices. Recent research has demonstrated notable advancements in stock market prediction through the application of machine learning methodologies. Specifically, techniques such as Support Vector Machine (SVM) and Random Forest (RF) [10] have gained prominence. Moreover, neural network-based approaches, including Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), as well as deep learning architectures like Long Short Term Memory (LSTM), have exhibited considerable promise in this regard [4] [11].

Artificial Neural Networks (ANNs) possess the capacity to uncover latent features through an autonomous learning process. They serve as adept approximators, adept at discerning complex input-output relationships within vast and intricate datasets. Consequently, ANNs emerge as a favorable choice for stock price prediction within organizational contexts. For instance, Selvin et al. conducted a predictive analysis of stock prices for companies listed on the National Stock Exchange (NSE) by employing a comparative assessment of various deep learning techniques [13]. In a similar vein, Hamzaebi et al. investigated the multi-periodic forecasting of stock markets through iterative and directive methodologies, employing ANNs as a primary model [14]. Additionally, Rout et al. applied a simplified Recurrent Neural Network (RNN) model to forecast stock market trends, conducting empirical evaluations on datasets from the Bombay Stock Exchange and the S&P 500 index [15].

Roman et al. conducted a study wherein Recurrent Neural Network (RNN) models were implemented on stock market data from five distinct countries: Canada, Hong Kong, Japan, the UK, and the USA. These models were trained on the respective datasets, and subsequently utilized to forecast trends in stock returns [17]. Similarly, in 2014, Yunus et al. employed Artificial Neural Networks (ANNs) on NASDAQ data to predict the closing prices of stocks [16]. Furthermore, Mizuno et al. utilized ANN for technical analysis on the Tokyo Stock Price Index (TOPIX) dataset, demonstrating its efficacy in developing a predictive system for timing buy and sell decisions [18]. Several studies have proposed the utilization of Random Forest (RF) for forecasting endeavors. RF, recognized as an ensemble technique, demonstrates proficiency in both regression and classification tasks. Its operational mechanism involves the construction of multiple decision trees during the training phase, with the aggregated output typically representing the mean regression of individual decision trees [19]. For instance, Mei et al. effectively applied RF to forecast real-time prices within the New York electricity market [20]. Similarly, Yand et al. conducted analogous research, employing the RF model for short-term load forecasting within the operational framework of electrical power systems [21]. Additionally, Herrera et al. utilized RF as a predictive model for the hourly forecasting of urban water demand [22].

In this study, two distinct methodologies, namely Artificial Neural Network (ANN) and Random Forest (RF), were employed to forecast the closing price of a given organization. These models were trained using a novel set of variables derived from the financial dataset, incorporating data points such as Open, High, Low, and Close prices specific to the company under consideration. These newly created indicators are anticipated to significantly enhance the accuracy of the models in predicting the subsequent day's closing price for the organization. The efficacy of these models was assessed through the utilization of two performance metrics: Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

The remainder of the papers is as follows: Section 2 provides methodology of techniques applied. Section 3 discusses the result with section 4 concluding the paper.

II. METHODOLOGY

2.1. Dataset description

The historical data for the five companies was gathered from Yahoo Finance [23]. This dataset spans a period of ten years, ranging from April 5, 2009, to April 5, 2019, encompassing the stock market activity of Nike, Goldman Sachs, Johnson & Johnson, Pfizer, and JP Morgan Chase & Co. The dataset comprises various attributes detailing the stock's performance, including High, Low, Open, Close, Adjusted Close, and Volume. For the purposes of this study, only the daily closing prices of the stocks were extracted. Table 2 presents the statistical summary of the dataset utilized for both training and testing phases.

2.2. Artificial neural network

Artificial Neural Networks (ANNs) stand out as an intelligent data mining technique adept at discerning underlying trends within datasets and extrapolating generalizations from them. Compared to conventional methods, ANNs exhibit prowess in simulating and analyzing intricate patterns present in unstructured data. The model is structured upon the foundational architecture of a Neural Network, comprising neurons organized into distinct layers. Typically, the model operates across three layers: the input layer, the hidden layer, and the output layer. The input layer incorporates newly engineered variables such as H-L (High-Low), O-C (Open-Close), 7-DAYS MA (Moving Average), 14-DAYS MA, 21-DAYS MA, 7-DAYS STD DEV (Standard Deviation), and Volume [23]. Each input's weight is multiplied, summed, and forwarded to the neurons within the hidden layer, also known as the activation layer. Here, the cumulative weights are computed and transmitted to the third layer, which constitutes the output layer. The output layer typically comprises a single neuron responsible for generating the predicted value, specifically the closing price of the stock. Figure 1 illustrates a comprehensive depiction of the ANN architecture, with the newly engineered variables serving as input to the model.

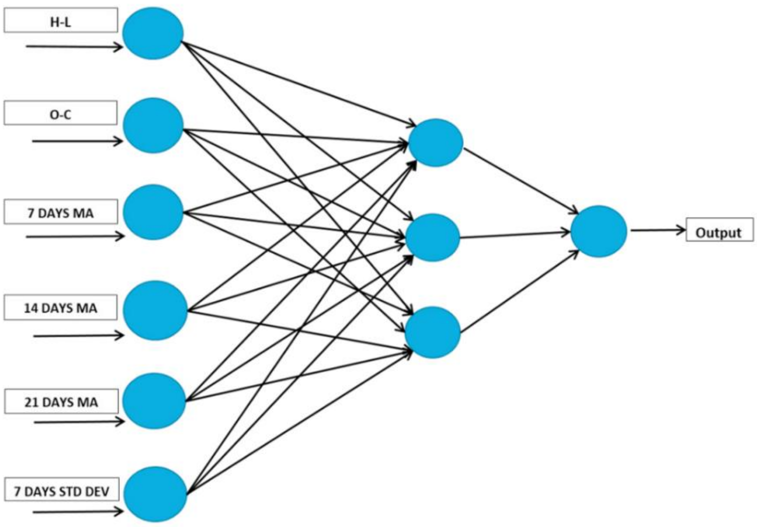

Fig 1: Proposed Artificial Neural Network Architecture

2.3. Random forest

Random Forest (RF) stands as an ensemble machine learning technique renowned for its capability to undertake both regression and classification tasks. The underlying concept revolves around amalgamating multiple decision trees to ascertain the final output, thereby mitigating reliance on individual decision trees and diminishing variance within the model. In the context of this study, newly engineered variables are furnished for the training of each decision tree, consequently influencing the decisions made at the nodes of the tree. Notably, the inherent noise prevalent in stock market data, owing to its extensive scale, often leads to significant fluctuations, potentially diverging the growth trajectory of the trees from the anticipated pattern. Hence, the primary objective is to minimize forecasting errors by treating stock market analysis as a classification problem, whereby the model utilizes training variables to predict the subsequent day's closing price of the stock for a specific company.

III. RESULTS

In order to assess the efficacy of the models, a comparative analysis is conducted between the two techniques across five distinct sector companies, namely, JP Morgan, Nike, Johnson & Johnson, Goldman Sachs, and Pfizer, employing both ANN and RF models. The predicted closing prices generated by these models are subjected to evaluation metrics including Root Mean

Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Bias Error (MBE). This rigorous evaluation process aims to identify and minimize the final errors in the predicted prices, thereby facilitating a comprehensive comparison of the predictive performance between the ANN and RF models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - F_i)^2}{n}} \qquad (1)$$

where 'Oi' refers to the original closing price, 'Fi' refers to the predicted closing price and 'n' refers to the total window size. MAPE has also been used to evaluate the performance of the model and is computed using eq. 2.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{(O_i - F_i)}{O_i} * 100 \qquad (2)$$

where 'Oi' refers to the original closing price, 'Fi' refers to the predicted closing price and 'n' refers to the total window size. MBE has also been used to evaluate the performance of the model and is computed using eq. 3.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(O_i - F_i) \qquad (3)$$

In the context of this study, 'Oi' denotes the original closing price of the stock, 'Fi' signifies the predicted closing price, and 'n' represents the total window size. Figure 2 illustrates graphs depicting the relationship between the original closing price of the stock and the predicted closing price for five different companies, employing the ANN model. Conversely, Figure 3 portrays similar graphs but utilizing the RF model. A comparative analysis of the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Bias Error (MBE) values obtained from both ANN and RF models is presented in Table 2. Notably, it is discerned that the ANN model yields superior prediction results for stock prices when compared to the RF model.

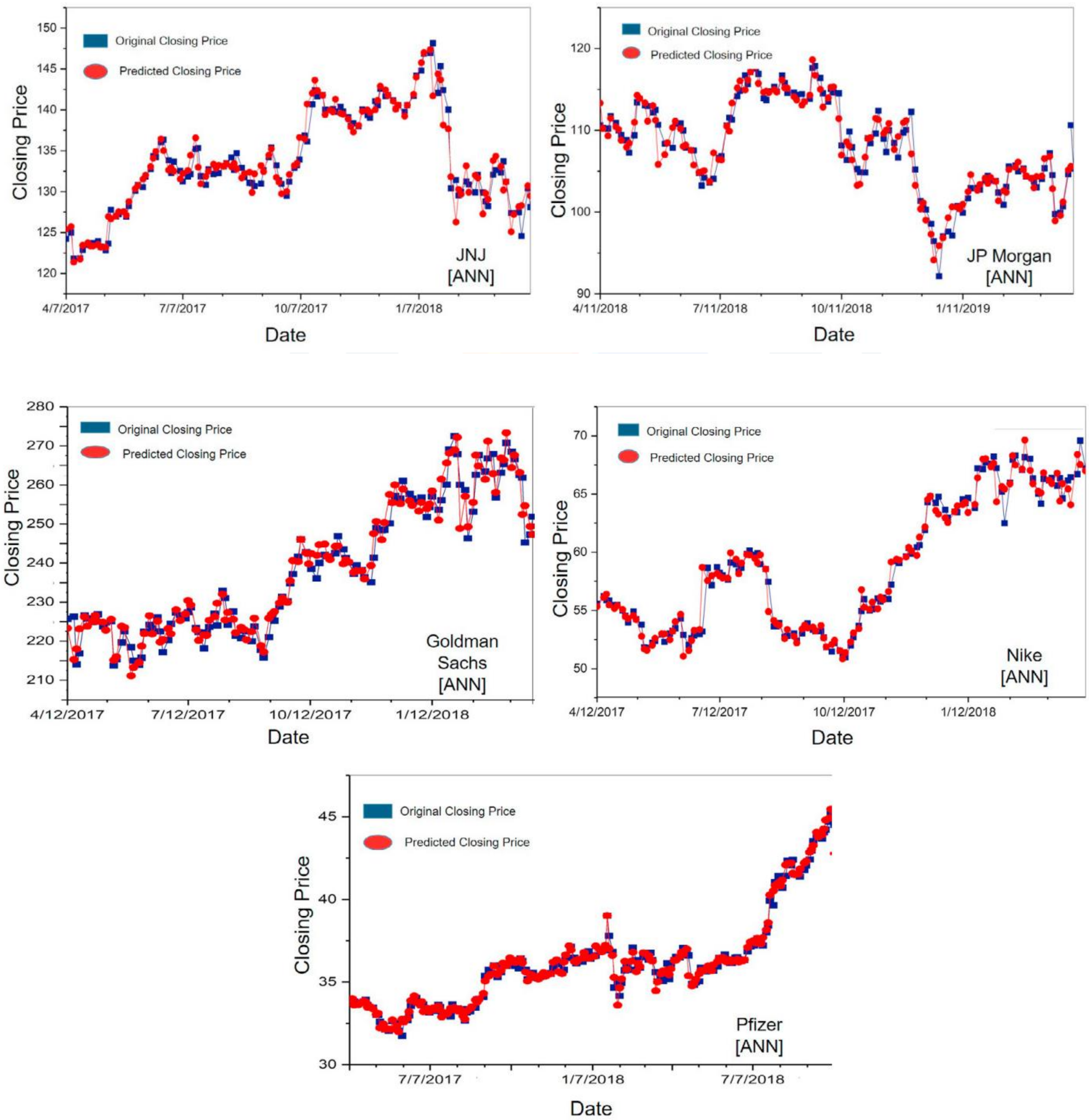

Fig 2: Predicted v/s original (expected) closing stock price using ANN.

IV. CONCLUSION

Forecasting stock market returns presents a formidable challenge, owing to the dynamic nature of stock values, which are influenced by a multitude of factors, thus forming intricate patterns. The historical dataset typically available on company websites often encompasses a limited set of features such as high, low, open, close, adjacent close values of stock prices, and trading volume, which may not suffice for accurate predictions. In pursuit of higher prediction accuracy, novel variables are derived from existing ones. In this study, the ANN model is employed to predict the next day's closing price of the stock, and for comparison, the

RF model is also implemented. A comparative analysis based on metrics such as RMSE, MAPE, and MBE unequivocally demonstrates the superior predictive performance of the ANN model over RF. Specifically, the best-performing ANN model yields RMSE (0.42), MAPE (0.77), and MBE (0.013) values. For future research endeavors, the exploration of deep learning models incorporating financial news articles alongside traditional financial parameters like closing price, trading volume, profit and loss statements, among others, holds promise for potentially enhancing prediction outcomes.

REFERENCES

[1] Masoud, Najeb MH. (2017) "The impact of stock market performance upon economic growth." International Journal of Economics and Financial Issues 3 (4) : 788–798.

[2] Murkute, Amod, and Tanuja Sarode. (2015) "Forecasting market price of stock using artificial neural network." International Journal of Computer Applications 124 (12) : 11-15.

[3] Hur, Jung, Manoj Raj, and Yohanes E. Riyanto. (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.

[4] Li, Lei, Yabin Wu, Yihang Ou, Qi Li, Yanquan Zhou, and Daoxin Chen. (2017) "Research on machine learning algorithms and feature extraction for time series." IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC): 1-5.

[5] Seber, George AF and Lee, Alan J. (2012) "Linear regression analysis." John Wiley & Sons 329

[6] Reichek, Nathaniel, and Richard B. Devereux. (1982) "Reliable estimation of peak left ventricular systolic pressure by M-mode echographicdetermined end-diastolic relative wall thickness: identification of severe valvular aortic stenosis in adult patients." American heart journal 103 (2) : 202-209.

[7] Chong, Terence Tai-Leung, and Wing-Kam Ng. (2008) "Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30." Applied Economics Letters 15 (14) : 1111-1114.

[8] Zhang, G. Peter. (2003) "Time series forecasting using a hybrid ARIMA and neural network mode." Neurocomputing 50 : 159-175.

[9] Suykens, Johan AK, and Joos Vandewalle. (1999) "Least squares support vector machine classifiers." Neural processing letters 9 (3) : 293-300.

[10] Liaw, Andy, and Matthew Wiener. (2002) "Classification and regression by Random Forest." R news 2 (3) : 18-22.

[11] Oyeyemi, Elijah O., Lee-Anne McKinnell, and Allon WV Poole. (2007) "Neural network-based prediction techniques for global modeling of M (3000) F2 ionospheric parameter." Advances in Space Research 39 (5) : 643-650.