Journal of Mathematics and Statistics Studies
ISSN: 2709-4200
DOI: 10.32996/jmss
Journal Homepage: www.al-kindipublisher.com/index.php/jmss

JMSS

AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Predicting S&P 500 Closing Prices Using a Feedforward Neural Network: A Machine Learning Approach

Hirushi Dilpriya[1]✉ Jayantha Lanel[2], Thamali Perera[3] and Chathuranga Vidanage[4]
[1]*Department of Computer and Data Science, Faculty of Computing, NSBM Green University, Pitipana-Thalagala Rd, Homagama 10200, Sri Lanka*
[234]*Department of Mathematics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Gangodawila, Nugegoda 10250, Sri Lanka*
Corresponding Author: Hirushi Dilpriya, **E-mail**: hirushi.d@nsbm.ac.lk

| ABSTRACT

The Standard and Poor's 500 index is a crucial benchmark for investors, financial analysts, and policymakers to assess stock market performance and make informed investment decisions. Accurate S&P 500 closing price prediction is vital for efficient portfolio management and risk mitigation. This study explores the application of neural networks, specifically the Feedforward Neural Network, to predict the S&P 500 closing price. The author used a dataset from 1950 to 2024, including variables such as open, high, low, close, and volume as the initial step. Various training-testing ratios are tested to evaluate the models' performance. The results highlight that an eighty to twenty split yields the best predictive accuracy, with the lowest Mean Absolute Error and Mean Absolute Percentage Error. Additionally, this study compares the Feedforward Neural Network's predictions to polynomial regression models and investigates cluster-wise fitting techniques to enhance accuracy. The findings demonstrate that neural networks can significantly improve the predictive power of S&P 500 closing prices, particularly when combined with advanced regression techniques, providing valuable insights for academic research and practical financial decisions.

| KEYWORDS

S&P 500 index, closing prices, Feedforward Neural Network, Machine Learning, New York, Testing Ratio

| ARTICLE INFORMATION

ACCEPTED: 01 February 2025 **PUBLISHED:** 26 February 2025 **DOI:** 10.32996/jmss.2025.6.1.3

1. Introduction

The S&P 500 index close price is a specific indicator in the stock market. It represents the collective performance of five hundred of the largest publicly traded companies in the US. It offers insights into the overall health of the stock market, shaping investment strategies, portfolio management, and broader economic assessments as a key benchmark. Investors, analysts, and policymakers continuously monitor the S&P 500 closing price. These observations assist in identifying market trends, assessing economic conditions, and making informed decisions about asset allocation and risk management. A notable rise in the S&P 500 index indicates strong corporate performance, promoting investor confidence and driving increased market participation. For example, the strong performance of the S&P 500 in 2019 led to an optimistic economic outlook, fueling higher investment levels. Conversely, a sharp decline in the index, such as during the COVID-19-generated crash in March 2020, signals market crashes and bubbles.

The dramatic drop in 2020 caused widespread panic and triggered significant selloffs, underscoring the index's role as a barometer of economic challenges.

Beyond US borders, the S&P 500's movements have global implications. A strong performance in the index often reflects confidence in the US economy, leading to positive ripple effects in international markets. On the flip side, steep declines can trigger global market selloffs as investors worldwide adjust to perceived risks in the US economy. Forecasting the S&P 500 close price is a complex task that involves a variety of methods, from traditional approaches like fundamental and technical analysis to advanced machine learning techniques. Fundamental analysis focuses on macroeconomic indicators such as GDP growth, unemployment, inflation, and interest rates, as well as corporate earnings reports and guidance. Strong economic data or positive earnings can drive the index upward, while poor indicators can lead to declines. Technical analysis uses historical price patterns, such as Moving Averages (MA), Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD), to predict future trends. Numerical analysis, which tracks changes in trading activity, is another useful tool for gauging market strength.

Quantitative models like linear regression and Auto-Regressive Integrated Moving Average (ARIMA) analyze historical data to predict future prices. Factor models consider variables like market risk, momentum, and company size, and they assist in making precise predictions. Sentiment analysis, which examines news and social media trends, offers additional insight into market sentiment. Tools like the Volatility Index measure market sentiment, providing context for investor behavior and potential price movements.

In recent years, machine learning and AI have significantly improved the accuracy of stock price predictions. Supervised learning models, including decision trees and neural networks, are trained on historical data to make predictions. Unsupervised learning, such as clustering algorithms, helps identify patterns that cannot easily be discernible through traditional methods. Reinforcement learning, which uses trial and error to optimize decisions, is increasingly applied to trading strategies. Supply and demand analyses, which examine corporate buybacks, IPOs, and economic policies, are essential components of financial models used for price prediction. Macroeconomic models that integrate various economic indicators also provide a broader perspective on market direction.

Although these diverse approaches improve prediction accuracy, financial markets remain inherently volatile and complex, leaving all forecasts subject to uncertainty. Addressing gaps in existing models and advancing research could lead to more reliable forecasting techniques.

2. Literature Review

The application of neural networks and machine learning in stock market prediction and financial forecasting has seen incredible advancements. Among these, Long Short-Term Memory (LSTM) networks are widely recognized because of their ability to process serial data effectively. Abed et al. (2021) developed LSTM networks to predict monthly pan evaporation, highlighting their efficacy in capturing temporal patterns in complicated datasets [1]. Similarly, Fischer and Krauss (2017) illustrated the potential of LSTM networks for financial market predictions, where their ability to learn long-term dependencies led to superior performance compared to traditional methods [4]. Nelson, Pereira, and De Oliveira (2017) also showcased LSTM's ability to predict stock market price movements accurately and emphasized the importance of feature engineering [15].

Di Persio and Honchar (2016) provided a comparative analysis of Artificial Neural Network (ANN) architectures for stock price prediction [2]. Also, they demonstrate their versatility in financial applications. Expanding on this, Selvin et al. (2017) implemented LSTM, Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN) in a sliding window framework for stock price predictions, with LSTM emerging as the most applicable model [18]. Meanwhile, Pang et al. (2018) proposed a unique neural network approach integrating high-order time series information, which improved predictive accuracy [16].

Beyond ANN and LSTM, other models have been explored for specific financial forecasting requirements. For instance, Zhang et al. (2019) employed Generative Adversarial Networks (GANs) to predict stock market trends, highlighting the ability of GANs to generate realistic man-made data and enhance model training. Guo et al. (2021) developed a deep neural network for predicting Jane Street market outcomes, emphasizing the significance of domain-specific datasets and techniques [5]. Hiransha et al. (2018) demonstrated the adaptability of deep learning models in predicting New York Stock Exchange stock trends, highlighting the flexibility of neural networks [10].

Interdisciplinary approaches also play a crucial role in financial forecasting. Kengatharan (2014) analyzed behavioral factors affecting investment decisions in the Colombo Stock Exchange, presenting insights into investor psychology. According to the policy analysis, Dilpriya, Lanel, and Perera (2023) reviewed Federal Reserve Bank policies using time series models to identify their impact on economic indicators [3]. Meanwhile, Hernandez et al. (2007) investigated accounting standard convergence for Latin American banks, providing a financial regulation perspective [6].

Comparative and systematic reviews offer a primary understanding of machine learning applications in stock markets. Mintarya et al. (2023) [12] conducted a systematic literature review of machine learning approaches, while Soni, Tewari, and Krishnan (2022) reviewed the advancements in stock price prediction techniques [21]. Vargas, De Lima, and Evsukoff (2017) explored deep learning models applied to financial news, highlighting the role of documented data in trend prediction [22]. Similarly, Maiti and Shetty (2020) [11] demonstrated the significance of deep learning in the Indian stock market, and Lamba et al. (2021) provided a comparative analysis of ANN for predicting Nifty 50 values for Indian Stock Market [8].

Additionally, initial studies like Moghaddam, and Esfandyari (2016) presented early evidence of ANN's utility in stock market index prediction [13]. Lichtner Bajjaoui (2021) offered a mathematical and statistical introduction to neural networks, enabling a deeper understanding of these systems [9]. Smith (1993) provided an earlier viewpoint on neural networks for statistical modeling, covering the way for their financial applications [20].

According to the reviewed literature, a clear gap exists between the available predictive models and the actual requirements of stock market analysis. Many studies focus on specific models, often improving them without conducting comparative analyses. Notably, eighty percent of the research utilized various neural network models for their predictions. However, researchers did not compare them to determine the most effective approach for forecasting the S&P 500 index. Anyhow, it is significant due to its role in the financial market. To address this gap, researchers aim to conduct comparative analyses to identify better prediction methods and enhance them through mathematical improvements, as existing models lack such enhancements. The authors begin their analysis part with the Feed Forward Neural Network (FFNN) in this article.

3. Materials and Methods

Among the various methods for predicting the close price of the S&P 500 Index, researchers identified neural networks as the most effective technique. As an initial step, the author reviewed various types of neural networks and selected FFNN as the first step. This model was chosen to determine the better neural network model for accurate predictions.

3.1 Feedforward Neural Network: FFNNs are primitive architecture in the ANNs. These networks are characterized by their unidirectional flow of information from input to output, making them essential in various applications, such as image recognition and natural language processing. This section explores the mathematical and statistical foundations underpinning FFNNs, elucidating their structure, training mechanisms, and theoretical principles. The structure of an FFNN includes multiple layers: an input layer, one or more hidden layers, and an output layer. Each layer consists of a certain number of neurons, or nodes, that are interconnected without forming cycles.

Mathematical Behavior: Consider an FFNN with L layers. The input to the network is denoted as $x \in R^d$, where d is the dimension of the input vector. Each layer l contains n_l neurons, and the output of the l^{th} layer is represented by the vector a^l. The transformation from one layer to the next is governed by the equation following equation 1:

$$a^{(l+1)} = f(W^l a^l + b^l) \tag{1}$$

where W^l is the weight matrix, b^l is the bias vector, and f is the activation function. Activation functions introduce non-linearity into the network, enabling it to approximate complex functions. Common activation functions include the sigmoid function $\sigma(z) = \frac{1}{1+e^{-z}}$, the hyperbolic tangent function $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$, and the Rectified Linear Unit (ReLU), $ReLU(z) = \max(0, z)$. FFNN includes multiple layers: an input layer, one or more hidden layers, and an output layer. Each layer consists of a certain number of neurons, or nodes, which are interconnected without forming cycles.

Training Mechanism: The training of FFNNs involves adjusting the weights and biases to minimize a loss function. This is typically done using a variant of the gradient descent algorithm.

Loss Function: The choice of loss function depends on the specific task. For regression tasks, the mean squared error (MSE) is often used:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2}$$

where y_i is the true value and $\hat{y}_i$ is the predicted value. For classification tasks, the cross-entropy loss is common:

$$Cross - Entropy = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij} log(\hat{y}_{ij}) \tag{3}$$

Training of FFNNs involves adjusting the weights and biases to minimize a loss function. This is typically done using a variant of the gradient descent algorithm. Backpropagation is the cornerstone algorithm for training FFNNs. It involves two main steps: Forward Pass: Compute the relevant output of the network for a given input. Backward Pass: Calculate the gradient of the loss function concerning each weight using the chain rule of calculus. The gradients are then used to update the weights via gradient descent. Mathematically, the weight update rule for gradient descent is given by:

$$W^{(l)} \leftarrow W^{(l)} - \eta \frac{\partial \mathcal{L}}{\partial W^{(l)}} \tag{4}$$

Where η is the learning rate and $\mathcal{L}$ is the loss function.

3.2 Application of Neural Network:

Based on the data from Yahoo Finance, the author selected date, open, high, low, close, and volume as variables for the neural networks. The next step was to identify the input and output variables. The common practice in the literature guided the researchers to use the mentioned variables as inputs, apart from the close price, which was chosen as the output variable. Once the author established the fundamental requirements for the neural network application, thought turned to the data. Initially, the researcher worked with 18,738 data points from 1950 to 2024. Subsequently, by applying cluster-wise linear fitting models, the sample size was adjusted to enhance accuracy and exclude outliers, thereby refining the results for the model. The basic neural network structure comprises five input variables and one output variable. For this analysis, the writer utilized two types of code. Firstly, the writer needed to determine the number of hidden layers and neurons per hidden layer. The first code defined FFNN using TensorFlow's Keras API. Initially, the network had two hidden layers, each with 64 neurons.

The next step was to determine the optimal training and testing ratio. Literature reviews typically recommend a 70% training and 30% testing split as a best practice. However, the author chose to reassess this ratio by exploring a range from 0.1 to 0.5 for the training proportion. The analysis was conducted separately for each of these training ratios. Following the discussed methodology, the researcher obtained the results outlined below from the analysis.

4. Results

The author began by using data from 1950 to 2024 and applied FFNNs to predict the next 30 days, starting from the last date in the dataset. When determining the prediction dates, the author chose to exclude public holidays and weekends. Below are the predicted values obtained through the FFNNs using different testing ratios.

Table 1: This Predicted closing values for the S&P 500 index, based on historical data from 1950 to 2024.

Date	Actual	Predicted 1	Predicted 2	Predicted 3	Predicted 4	Predicted 5
		Test 0.5	Test 0.4	Test 0.3	Test 0.2	Test 0.1
2024.06.24	5447.87	4038.86	4214.47	4238.18	4399.14	4314.28
2024.06.25	5469.3	5127.32	5133.22	5096.08	4899.31	4837.70
2024.06.26	5477.9	4113.60	4237.66	4058.46	4108.36	3965.62
2024.06.27	5482.87	4521.78	4703.02	4740.41	4776.56	4672.18
2024.06.28	5460.48	4778.39	4774.64	4628.55	4473.39	4357.10
2024.07.01	5475.09	4511.07	4631.06	4439.40	4432.43	4252.54
2024.07.02	5509.01	4343.91	4628.66	4542.02	4614.08	4441.06
2024.07.03	5537.02	4673.86	4770.90	4655.76	4571.10	4397.22
2024.07.05	5567.19	4482.39	4731.67	4629.82	4625.17	4434.58
2024.07.08	5572.85	4541.83	4774.68	4626.32	4603.80	4392.41

2024.07.09	5576.98	4504.19	4757.82	4549.53	4584.48	4361.51
2024.07.10	5633.91	4443.01	4770.59	4592.21	4620.85	4393.549
2024.07.11	5584.54	4541.28	4800.05	4603.37	4604.59	4373.13
2024.07.12	5615.35	4466.48	4792.58	4571.29	4617.78	4377.33
2024.07.15	5631.22	4461.33	4833.79	4598.91	4638.43	4381.14
2024.07.16	5667.2	4498.79	4853.50	4616.94	4638.57	4376.91
2024.07.17	5588.27	4492.10	4861.26	4602.96	4644.69	4376.05
2024.07.18	5544.59	4470.31	4874.01	4618.16	4652.88	4378.77
2024.07.19	5505	4498.90	4888.54	4619.88	4654.69	4375.20
2024.07.22	5564.41	4490.99	4924.16	4627.48	4672.41	4375.22
2024.07.23	5555.74	4474.33	4935.67	4632.88	4679.08	4375.99
2024.07.24	5427.13	4483.53	4949.06	4638.52	4683.29	4374.80
2024.07.25	5399.22	4482.50	4960.71	4638.68	4689.49	4374.91
2024.07.26	5459.1	4474.70	4973.19	4645.33	4694.78	4374.66
2024.07.29	5463.54	4475.61	5010.45	4656.22	4710.82	4373.58
2024.07.30	5436.44	4478.85	5022.78	4658.51	4716.40	4373.34
2024.07.31	5522.3	4473.48	5035.07	4662.92	4721.83	4373.10
2024.08.01	5446.68	4475.23	5047.62	4666.79	4727.07	4372.70
2024.08.02	5346.56	4474.81	5059.92	4669.89	4732.61	4372.49
2024.08.05	5186.33	4471.40	5097.18	4681.42	4748.66	4371.59

Table 1, shows the actual closing values and the predicted values for each test ratio. Predicted Set 1 was obtained using a 0.5 test ratio, Predicted Set 2 was obtained using a 0.4 test ratio, and this pattern continues down to Predicted Set 5, which is based on a 0.1 test ratio. Researchers then sought to identify errors and assess the accuracy of the predicted values which are described in Table 1. Table 2 below provides details on the accuracy of these FFNN predictions.

Table 2: Errors for FFNN Predictions: 1950-2024.

	Predicted 1	**Predicted 2**	**Predicted 3**	**Predicted 4**	**Predicted 5**
	Test: 0.5	Test: 0.4	Test: 0.3	Test: 0.2	Test: 0.1
Test MAE	0.000676242	0.0011283	0.001071	0.00070334	0.0005856
MAPE	1.64%	1.24%	3.58%	1.06%	0.86%

According to Table 2, the lowest Mean Absolute Error (MAE) is observed with the fifth set of predictions, which corresponds to using 90 percent of the data for training and 10 percent for testing. Additionally, the model with the best Mean Absolute Percentage Error (MAPE) is the same. As the next step, the author wanted a model to predict the actual and forecasted closing values. They started by applying a linear fitted model and then evaluated the Root Mean Square Error (RMSE) for these models, as illustrated in Figure 1.

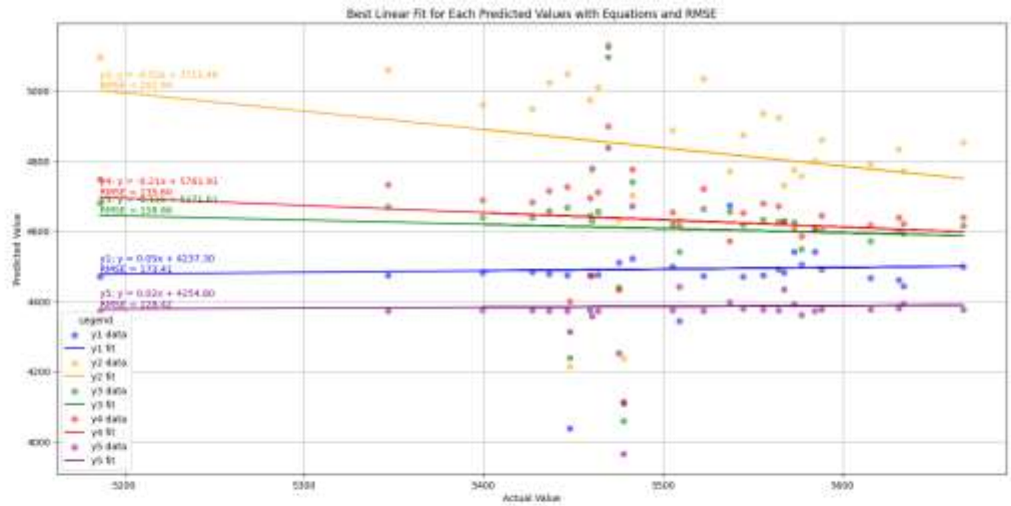

Figure 1: Relationship between Actual and Predicted closing values with the Fitted Line Equations 1950- 2024.

The most accurate illustration, with the lowest RMSE, is provided by the y5 data, which used prediction set 5 with a 0.1 training-testing ratio. This result aligns with the findings from the FFNN analysis. However, the fitted line models still exhibit comparatively high RMSE values. Consequently, the focus shifted to the best regression model rather than the fitted line equations. The results obtained from the regression models are as follows in Figure 2.

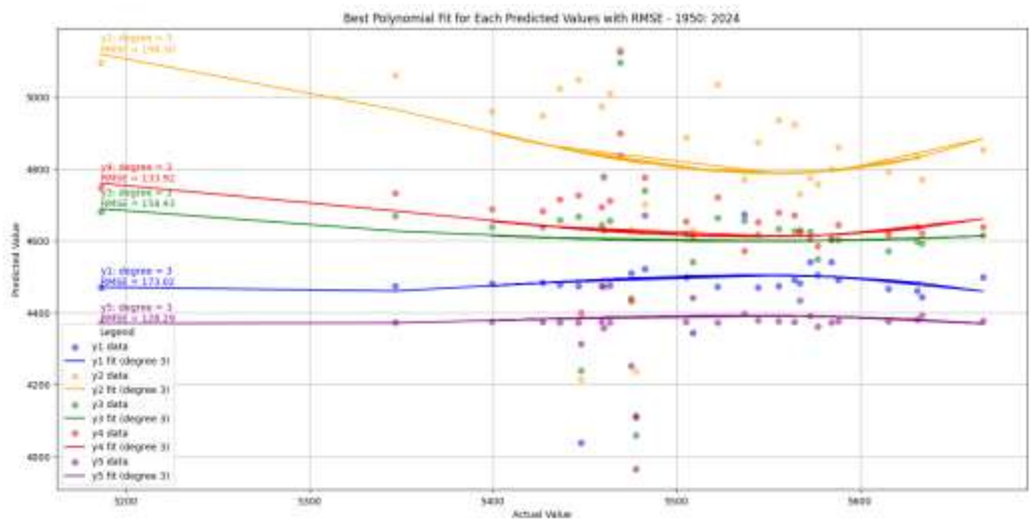

Figure 2: The Best Polynomial Fit for Each Predicted Value with RMSE 1950- 2024.

According to Figure 2, the best polynomial fit was identified at prediction set 5. This degree 3 polynomial model is less than 0.13 better than the fitted linear model. Nevertheless, among the models described so far, the training set 0.9 yielded the best results. Next, the researcher focused on addressing outliers and determining the optimal number of clusters to enhance prediction accuracy by minimizing MAE and MAPE through the FFNN.

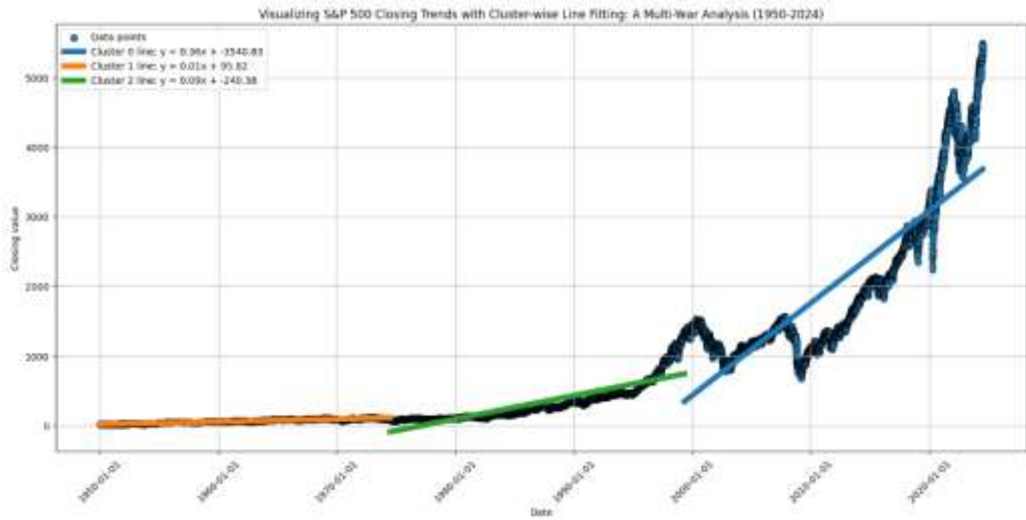

Figure 3: Visualizing S&P 500 Closing Trends with Cluster-wise Line Fitting.

According to Figure 3, there are three main clusters observed from 1950 to 2024. The most recent cluster, highlighted in blue, spans from 2000 to 2024. Consequently, researchers applied the same FFNNs to this period. The results for this cluster are detailed in Table 3 below.

Table 3: Predicted closing values for the S&P 500 index, based on historical data from 2000 to 2024.

Date	Actual	Predicted 1	Predicted 2	Predicted 3	Predicted 4	Predicted 5
		Test 0.5	Test 0.4	Test 0.3	Test 0.2	Test 0.1
2024.06.24	5447.87	4736.15	4479.84	4361.92	4035.75	4254.10
2024.06.25	5469.3	4865.28	4898.32	4955.29	5017.07	5007.08
2024.06.26	5477.9	4564.75	4474.44	4238.70	3557.52	3876.74
2024.06.27	5482.87	4930.34	4682.43	4669.96	4607.24	4731.14
2024.06.28	5460.48	4750.59	4677.36	4587.65	4315.46	4464.28
2024.07.01	5475.09	4883.52	4636.25	4456.49	3771.97	4205.21
2024.07.02	5509.01	4982.70	4634.91	4505.99	4096.22	4446.36
2024.07.03	5537.02	4976.31	4656.07	4543.40	4316.85	4515.55
2024.07.05	5567.19	5065.54	4648.64	4518.01	4302.29	4537.31
2024.07.08	5572.85	5164.29	4653.24	4494.20	4203.78	4493.36
2024.07.09	5576.98	5203.23	4653.39	4466.40	3682.20	4345.05
2024.07.10	5633.91	5240.76	4654.40	4472.84	4010.75	4454.64
2024.07.11	5584.54	5274.92	4656.82	4469.96	3944.15	4430.809
2024.07.12	5615.35	5314.46	4657.49	4457.74	3696.66	4387.898
2024.07.15	5631.22	5426.36	4661.84	4448.44	3805.43	4432.953
2024.07.16	5667.2	5463.97	4663.40	4446.48	3902.25	4449.671
2024.07.17	5588.27	5502.87	4664.76	4439.61	3641.37	4414.83
2024.07.18	5544.59	5541.25	4666.11	4436.79	3861.60	4453.881
2024.07.19	5505	5580.32	4667.55	4432.38	3725.69	4436.69
2024.07.22	5564.41	5698.96	4671.66	4419.25	3610.45	4438.47
2024.07.23	5555.74	5739.02	4673.02	4415.12	3711.32	4453.82
2024.07.24	5427.13	5779.44	4674.39	4411.10	3688.97	4454.18
2024.07.25	5399.22	5820.15	4675.75	4406.63	3580.39	4449.40
2024.07.26	5459.1	5861.13	4677.1	4402.62	3696.68	4462.13

2024.07.29	5463.54	5985.93	4681.14	4390.11	3622.44	4466.12
2024.07.30	5436.44	6028.14	4682.48	4385.91	3515.0	4463.22
2024.07.31	5522.3	6070.65	4683.82	4381.80	3585.18	4470.22
2024.08.01	5446.68	6113.47	4685.15	4377.69	3529.32	4470.29
2024.08.02	5346.56	6156.60	4686.47	4373.56	3495.61	4471.92
2024.08.05	5186.33	6287.90	4690.4	4361.32	3483.18	4480.10

Table 3, shows the actual closing values and the predicted values for each test ratio. Predicted set 1 was obtained using a 0.5 test ratio, predicted set 2 was obtained using a 0.4 test ratio, and this pattern continues down to predicted set 5, which is based on a 0.1 test ratio. The researcher then sought to identify errors and assess the accuracy of the predicted values which are described in Table 3. Table 4 below provides details on the accuracy of these FFNN predictions that are based on the data from 2000 to 2024.

Table 4: Errors for FFNN Predictions: 2000-2024.

	Predicted 1	**Predicted 2**	**Predicted 3**	**Predicted 4**	**Predicted 5**
	Test: 0.5	Test: 0.4	Test: 0.3	Test: 0.2	Test: 0.1
Test MAE	0.004602659	0.0018758	0.002026235	0.003480874	0.001427
MAPE	1.02%	0.51%	0.55%	0.85%	0.38%

According to Table 4, the lowest MAE is observed with the fifth set of predictions, which corresponds to using 90 percent of the data for testing and 10 percent for training. Additionally, the model with the best MAPE is the same. In the initial analysis, the model achieved a minimum MAPE of 0.86%. However, during the cluster-wise analysis, the MAPE further decreased, indicating an improvement in the model's performance. The researcher then attempted to identify the most suitable regression model for comparing actual closing values to predicted closing values. Figure 4 illustrates the selected regression model and presents the RMSE of the models from 2000 to 2024.

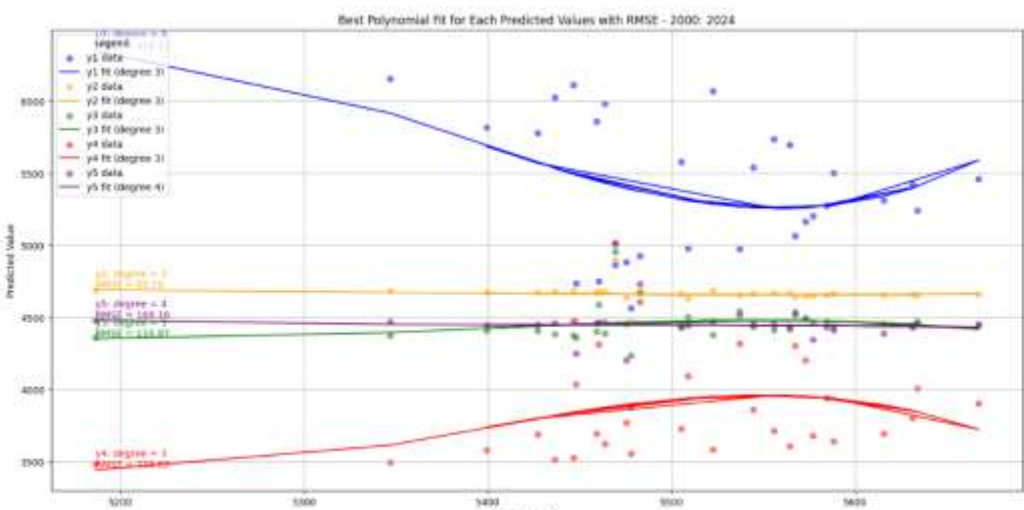

Figure 4: Best Linear Fit for Each Predicted Value with Equations and RMSE 2000 – 2024.

As shown in Figure 4, the model achieves a lower RMSE value compared to the best polynomial fit for the period from 1950 to 2024. Additionally, Figure 4 indicates that the best linear fit model is provided by Model 2, which was trained on 60% of the data and tested on the remaining 40%. The specified model produced an RMSE of 65.81.

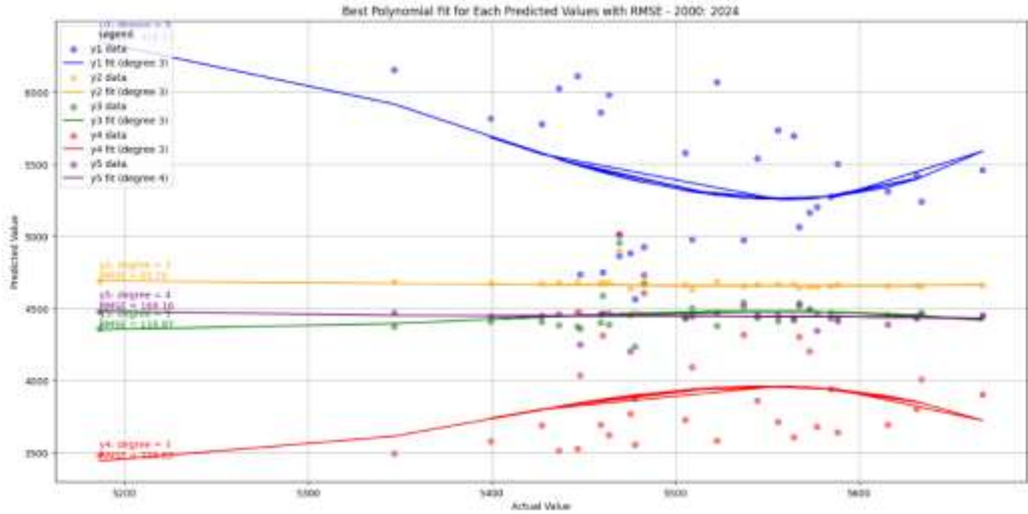

Figure 5: Best Polynomial Fit Each Predicted Value with RMSE 2000 – 2024.

As shown in Figure 5, the best polynomial model is provided by model 2 with 60 percent of the training data, highlighted in yellow. The RMSE value of this model is less than 0.11, outperforming the fitted linear model. To examine the fluctuations of each model in comparison to the actual closing values of the S&P 500 index, Figure 6 illustrates the performance of each model individually.

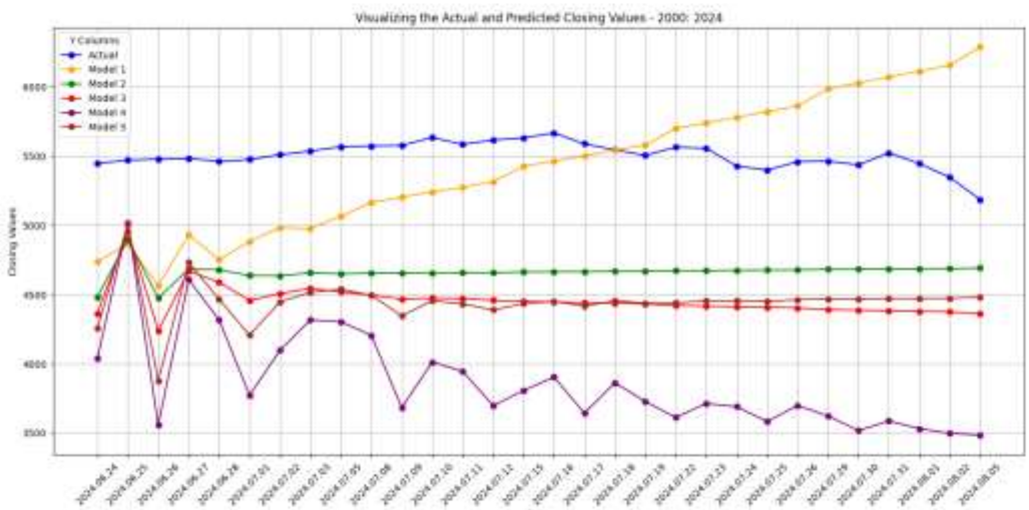

Figure 6: Visualizing the Actual and Predicted Closing Values 2000 – 2024.

Figure 6 illustrates the predicted values across various training-to-testing ratios. Model 4 displays characteristics that are closely aligned with the actual values. However, Models 2 and 3 provide predictions that are the closest to the actual values. The researcher aimed to determine the best-fitting line within the timeframe of 2000 to 2024. During this period, three main clusters were identified as follows in Figure 7.

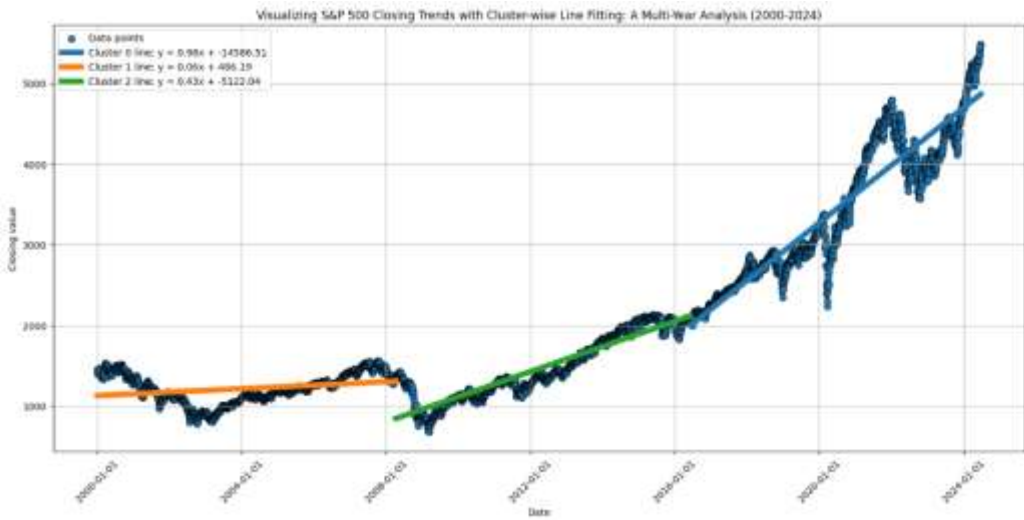

Figure 7: Visualizing S&P 500 Closing Trends with Cluster-wise Line Fitting 2000 – 2024.

As shown in Figure 7, three main clusters are identifiable. The authors selected the third cluster, spanning from 2016 to 2024, as it represents the most recent data. This cluster is highlighted by the blue fitted line.

Table 5: Predicted closing values for the S&P 500 index, based on historical data from 2016 to 2024.

Date	Actual	Predicted 1	Predicted 2	Predicted 3	Predicted 4	Predicted 5
		Test 0.5	Test 0.4	Test 0.3	Test 0.2	Test 0.1
2024.06.24	5447.87	5093.19	5139.21	4994.87	5137.98	5093.83
2024.06.25	5469.3	5151.30	5157.72	5427.79	5328.40	5263.29
2024.06.26	5477.9	5134.11	5094.42	4891.71	5106.82	4771.63
2024.06.27	5482.87	5087.34	5116.93	5186.21	5249.12	5280.24
2024.06.28	5460.48	5097.59	5089.89	5281.74	5229.95	4921.96
2024.07.01	5475.09	5054.61	5051.05	5052.88	5202.01	4722.65
2024.07.02	5509.01	5038.44	5042.07	4945.70	5220.45	5148.33
2024.07.03	5537.02	5024.72	5029.45	5302.51	5246.15	4893.27
2024.07.05	5567.19	4999.88	5004.69	5092.30	5244.15	5106.95
2024.07.08	5572.85	4962.00	4967.41	5208.49	5252.45	4874.91
2024.07.09	5576.98	4949.58	4954.93	4986.29	5244.23	4750.79
2024.07.10	5633.91	4937.11	4942.99	4883.98	5252.01	5051.92
2024.07.11	5584.54	4924.45	4930.87	5201.02	5257.86	4646.09
2024.07.12	5615.35	4912.08	4918.81	4819.05	5255.72	4943.46
2024.07.15	5631.22	4874.95	4883.06	4760.21	5268.29	5005.63
2024.07.16	5667.2	4862.66	4871.23	5112.14	5273.99	4634.51
2024.07.17	5588.27	4850.41	4859.44	4919.50	5275.45	4863.52
2024.07.18	5544.59	4838.23	4847.72	4820.69	5280.02	4861.76
2024.07.19	5505	4826.07	4836.03	5110.95	5283.75	4613.28
2024.07.22	5564.41	4789.87	4801.27	5016.05	5294.14	4784.42
2024.07.23	5555.74	4777.90	4789.78	4708.18	5297.77	4852.85

2024.07.24	5427.13	4765.96	4778.33	5026.94	5301.74	4561.03
2024.07.25	5399.22	4754.07	4766.93	4856.69	5305.06	4903.09
2024.07.26	5459.1	4742.22	4755.58	4759.04	5308.97	4632.43
2024.07.29	5463.54	4707.04	4721.81	4861.56	5320.18	4513.29
2024.07.30	5436.44	4695.49	4710.64	4945.71	5323.86	4841.87
2024.07.31	5522.3	4684.03	4699.52	4655.78	5327.66	4636.94
2024.08.01	5446.68	4672.67	4688.45	4951.04	5331.51	4619.22
2024.08.02	5346.56	4661.39	4677.42	4797.98	5335.27	4822.05
2024.08.05	5186.33	4628.08	4644.61	4663.42	5346.87	4650.81

Table 5 presents the predicted closing values of the S&P 500 index based on historical data from 2016 to 2024. The researcher analyzed the summarized data in this table, following the same approach used in previous cases.

Table 6: Errors for FFNN Predictions: 2016-2024.

	Predicted 1	Predicted 2	Predicted 3	Predicted 4	Predicted 5
	Test: 0.5	Test: 0.4	Test: 0.3	Test: 0.2	Test: 0.1
Test MAE	0.010497236	0.0109291	0.008229873	0.009964728	0.0079766
MAPE	0.47%	0.49%	0.37%	0.44%	0.36%

Based on Table 6, the lowest MAE was achieved with prediction model 5, which uses a 90% testing and 10% training split. Both prediction models 3 and 5 demonstrate better accuracy compared to the other models. Consequently, the analyst decided to evaluate the linear and polynomial best-fit models, as was done in previous cases.

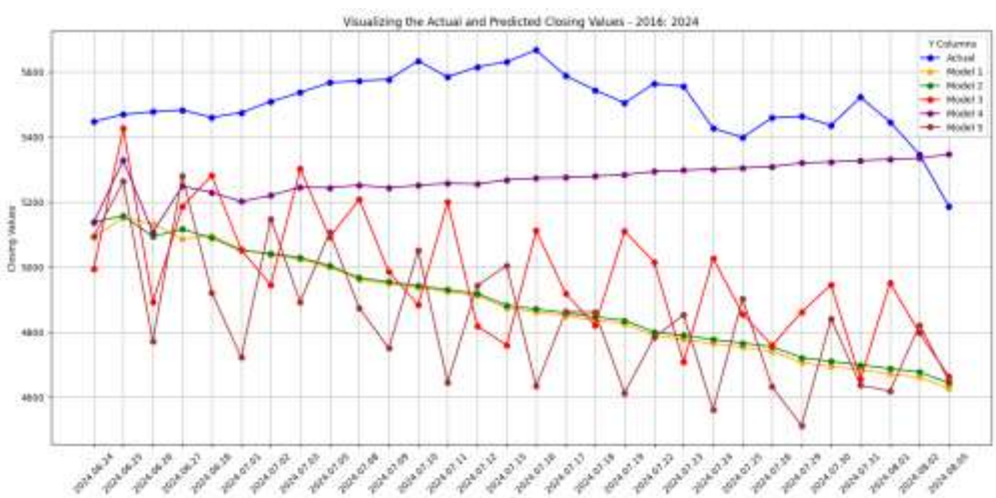

Figure 8: Best Linear Fit for Each Predicted Value with Equations and RMSE 2016 – 2024.

As shown in Figure 8, the model achieves a lower RMSE value compared to the best polynomial fit for the period from 2000 to 2024. Additionally, Figure 8 indicates that the best linear fit model is provided by Model 4, which was trained on 80% of the data and tested on the remaining 20%. The specified model produced an RMSE of 51.86 for model 4.

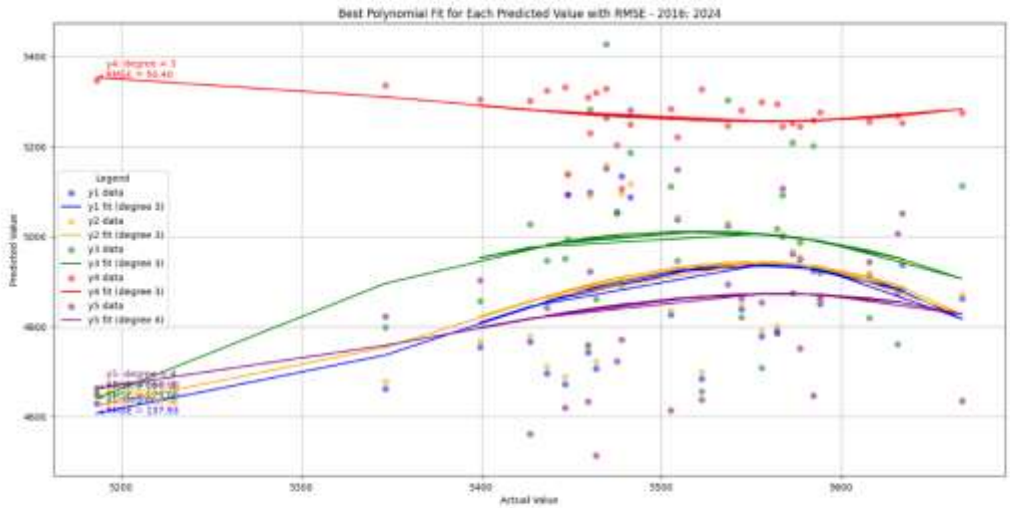

Figure 9. Best Polynomial Fit Each Predicted Value with RMSE 2016 – 2024.

As shown in Figure 9, the best polynomial model is provided by Model 4 with 80 percent of the training data, highlighted in red color. The RMSE value of this model is less than 1.46, outperforming the fitted linear model. To examine the fluctuations of each model in comparison to the actual closing values of the S&P 500 index, Figure 10 illustrates the performance of each model individually.

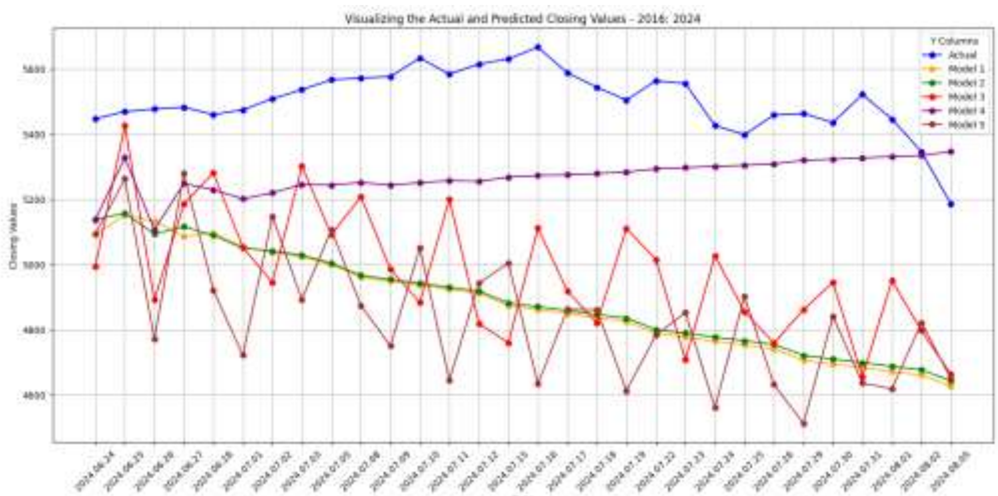

Figure 10. Visualizing the Actual and Predicted Closing Values 2016 – 2024.

According to Figure 10, the FFNN model closely approximates the actual closing values. However, Model 3 and Model 4 outperform the other models, with Model 4 providing the most accurate predictions, closely matching the actual values. Model 4 was built using an 80 percent training and 20 percent testing split. Furthermore, the lowest MSE and MAPE values were achieved with an 8:2 and 9:1 training-to-testing ratio. Thus, both the accuracy metrics and the graphical interpretations indicate consistent results.

5. Discussion

According to the literature review, most studies utilize machine learning and neural network concepts to predict stock market indices. However, it is challenging to find comparative studies or mathematical contributions in this domain. Therefore, the authors aim to provide a mathematical contribution to the early prediction of the S&P 500 index close price. As the first step, the authors sought to identify suitable models for these predictions. Approximately 80% of the reviewed articles highlighted neural networks as the most effective approach for such predictions. Consequently, the authors decided to begin their analysis using FFNN, and the results of this analysis are presented in this article.

Initially, there was some uncertainty regarding the appropriate training-to-testing ratio for the analysis. Most literature suggests a 0.7 training-to-testing ratio as optimal. Guided by this, the author experimented with ratios ranging from 0.5 to 0.9 for the dataset. The literature also commonly indicates that predictions over short time horizons, such as 5 to 10 days, yield better results in similar cases. However, this analysis revealed that the highest accuracy was achieved for 30-day predictions. Despite these findings, the author recognized challenges in pinpointing the optimal time for accurate results due to significant fluctuations in the S&P 500 closing prices caused by various financial decisions and crises over time. To address this, the author aimed to identify the most suitable timeframes for this research by analyzing the complete dataset from 1950 to 2024, ensuring outliers were excluded.

The initial prediction was conducted on the entire dataset, achieving the best results with a 0.9 training-to-testing ratio, characterized by low values of MAE and MAPE. Following this, regression analysis was applied to the corresponding predictions to identify the best-fitted line. While the most accurate model was found using a 0.9 training-to-testing ratio, followed by 0.8, the MSE values were higher than expected.

To address this, researchers segmented the dataset based on its characteristics, dividing it into multiple clusters and focusing on recent data. The clustering process was repeated iteratively, refining the groups according to their distinct features. This approach led to a noticeable improvement in accuracy, as indicated by reduced MAE and MAPE values. However, the researchers observed a turning point where continued splitting resulted in increasing errors.

Ultimately, the best-performing model was identified for data spanning the period from 2016 to 2024, with a 0.8 training-to-testing ratio. This analysis helped determine the key characteristics required for more accurate predictions using FFNN. By comparing the predicted and actual values, the author recognized discrepancies and the need for mathematical enhancements to improve the model. Before implementing these mathematical contributions, it is essential to identify a comparatively better Neural Network model for future research.

6. Study Limitations and Future Research

At the beginning of this research, the author aimed to analyze data fluctuations on an hourly basis for a specific day. However, due to limited data availability, the study relied on daily data instead. This limitation had a minimal impact on the results, as the analysis was based on average daily values. Nevertheless, using hourly data could have reduced errors compared to the current findings. The results were obtained using a Python program, and each network analysis was conducted based on standard neural network concepts. However, further improvements could be achieved by incorporating advanced theoretical aspects of neural networks. The author utilized only the S&P 500 index for this analysis. However, future implementations could incorporate other indices to expand the scope of the study. Additionally, while the researcher initially employed a feedforward neural network for fundamental analysis, similar approaches can be applied to other types of neural networks. This would help identify the most effective model while contributing valuable mathematical and statistical insights. Based on the results of this analysis, future investors and limited companies can determine their next steps in the financial markets.

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] Abed, Mustafa, Monzur Alam Imteaz, Ali Najah Ahmed, and Yuk Feng Huang. (October 20, 2021). "Application of Long Short-term Memory Neural Network Technique for Predicting Monthly Pan Evaporation." *Scientific Reports* 11, no. 1 https://doi.org/10.1038/s41598-021-99999-y

[2] Di Persio, Luca, and Oleksandr Honchar. (January 1, 2016). "Artificial Neural Networks architectures for stock price prediction: comparisons and applications." *INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING* 10: 403–13. https://iris.univr.it/handle/11562/955101

[3] Dilpriya T. A. H, Lanel G. H. J, and Perera M. T. M. (January 1, 2023). "Reviewing the Efficacy of Federal Reserve Bank Reserve Policies Through a Time Series Analysis of the Effective Federal Funds Rate." *International Journal of Research and Innovation in Social Science* VII, no. IV: 869–80. https://doi.org/10.47772/ijriss.2023.7472

[4] Fischer, Thomas, and Christopher Krauss. (December 6, 2017). "Deep Learning with Long Short-term Memory Networks for Financial Market Predictions." *European Journal of Operational Research* 270, no. 2: 654–69. https://doi.org/10.1016/j.ejor.2017.11.054.

[5] Guo, Shuting, Haoshan Wang, Lin Jiahao, and Xi Chen. (2021). *A Deep Neural Network Based Model for Jane Street Market Prediction*. *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE 2021)*. https://doi.org/10.1109/icbaie52039.2021.9390063

[6] Hernandez, Salvador Marin, Mercedes Palacios Manzano, Alejandro Hazera, and Carmen Quirvan. (January 1, 2007). "Latin American Banking Institutions Trading on NewYork Stock Exchange: Convergence–Divergence of Latin American Accounting Standards and US GAAP." *Advances in International Accounting* 20: 225–51. https://doi.org/10.1016/s0897-3660(07)20008-x

[7] Kengatharan, Lingesiya, and Navaneethakrishnan Kengatharan. (February 22, 2014). "The Influence of Behavioral Factors in Making Investment Decisions and Performance: Study on Investors of Colombo Stock Exchange, Sri Lanka." *Asian Journal of Finance & Accounting* 6, no. 1: 1. https://doi.org/10.5296/ajfa.v6i1.4893

[8] Lamba, Vikas, Susheela Hooda, Vikas Solanki, Rakesh Ahuja, Sachin Ahuja, and Amandeep Kaur. (October 22, 2021). "Comparative Analysis of Artificial Neural Networks for Predicting Nifty 50 Value in the Indian Stock Market." *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*). https://doi.org/10.1109/iscon52037.2021.9702400

[9] Lichtner-Bajjaoui, Aisha. (2021). "A Mathematical Introduction to Neural Networks." http://diposit.ub.edu/dspace/bitstream/2445/180441/2/tfm_lichtner_bajjaoui_aisha.pdf

[10] M, Hiransha, Gopalakrishnan EA, Vijay Krishna Menon, and Soman KP. (January 1, 2018). "NSE Stock Market Prediction Using Deep-Learning Models." *Procedia Computer Science* 132: 1351–62. https://doi.org/10.1016/j.procs.2018.05.050

[11] Maiti, Ayan, and Pushparaj Shetty D. (2020). *Indian Stock Market Prediction using Deep Learning*. *2020 IEEE REGION 10 CONFERENCE (TENCON)*. https://doi.org/10.1109/tencon50793.2020.9293712

[12] Mintarya, Latrisha N., Jeta N.M. Halim, Callista Angie, Said Achmad, and Aditya Kurniawan. (January 1, 2023). "Machine Learning Approaches in Stock Market Prediction: A Systematic Literature Review." *Procedia Computer Science* 216: 96–102. https://doi.org/10.1016/j.procs.2022.12.115

[13] Moghaddam, Amin Hedayati, Moein Hedayati Moghaddam, and Morteza Esfandyari. (July 29, 2016). "Stock Market Index Prediction Using Artificial Neural Network." *Journal of Economics Finance and Administrative Science* 21, no. 41: 89–93. https://doi.org/10.1016/j.jefas.2016.07.002

[14] Narayan, Paresh Kumar, and Huson Ali Ahmed. (January 1, 2014). "Importance of Skewness in Decision Making: Evidence from the Indian Stock Exchange." *Global Finance Journal* 25, no. 3: 260–69. https://doi.org/10.1016/j.gfj.2014.10.006

[15] Nelson, David M. Q., Adriano C. M. Pereira, and Renato A. De Oliveira. (May 1, 2017). "Stock Market's Price Movement Prediction with LSTM Neural Networks." *2022 International Joint Conference on Neural Networks (IJCNN)*, 1419–26. https://doi.org/10.1109/ijcnn.2017.7966019

[16] Pang, Xiongwen, Yanqiang Zhou, Pan Wang, Weiwei Lin, and Victor Chang. (January 12, 2018). "An innovative neural network approach for stock market prediction." *The Journal of Supercomputing* 76, no. 3: 2098–2118. https://doi.org/10.1007/s11227-017-2228-y

[17] Qiu, Mingyue, and Yu Song. (May 19, 2016). "Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model." *PLoS ONE* 11, no. 5: e0155133. https://doi.org/10.1371/journal.pone.0155133

[18] Selvin, Sreelekshmy, R Vinayakumar, E. A Gopalakrishnan, Vijay Krishna Menon, and K. P. Soman. (September 1, 2017). "Stock price prediction using LSTM, RNN and CNN-sliding window model." *ICACCI.2017.8126078*, 1643–47. https://doi.org/10.1109/icacci.2017.8126078

[19] Shen, Jingyi, and M. Omair Shafiq. (August 28, 2020). "Short-term Stock Market Price Trend Prediction Using a Comprehensive Deep Learning System." *Journal of Big Data* 7, no. 1. https://doi.org/10.1186/s40537-020-00333-6

[20] Smith, Murray. (1993). *Neural Networks for Statistical Modeling.* http://ci.nii.ac.jp/ncid/BA29818777

[21] Soni, Payal, Yogya Tewari, and Deepa Krishnan. (January 1, 2022). "Machine Learning Approaches in Stock Price Prediction: A Systematic Review." *Journal of Physics Conference Series* 2161, no. 1: 012065. https://doi.org/10.1088/1742-6596/2161/1/012065

[22] Vargas, Manuel R., Beatriz S. L. P. De Lima, and Alexandre G. Evsukoff. (2017). *Deep learning for stock market prediction from financial news articles*. https://doi.org/10.1109/civemsa.2017.7995302

[23] Wen, Min, Ping Li, Lingfei Zhang, and Yan Chen. "Stock Market Trend Prediction Using High-Order Information of Time Series." *IEEE Access* 7, January 1, 2019: 28299–308. https://doi.org/10.1109/access.2019.2901842.

[24] Wijesundera, A. a. V. I., D. a. S. Weerasinghe, T. P. C. R. Krishna, M. M. D. Gunawardena, and H. R. I. Peiris. "Predictability of Stock Returns Using Financial Ratios: Empirical Evidence from Colombo Stock Exchange." *Deleted Journal* 4, no. 2, September 6, 2016: 44–55. https://doi.org/10.4038/kjm.v4i2.7500.

[25] Zhang, Kang, Guoqiang Zhong, Junyu Dong, Shengke Wang, and Yong Wang. "Stock Market Prediction Based on Generative Adversarial Network." *Procedia Computer Science* 147, January 1, 2019: 400–406. https://doi.org/10.1016/j.procs.2019.01.256.