Pakistan Journal of Law, Analysis and Wisdom
Volume No. 3, Issue No. 1, January 2024
e-ISSN: 2959-0825, p-ISSN: 2959-0817
http://pjlaw.com.pk

Stock Price Prediction Using Neural Networks

Sakina
Ph.D. Scholar, Department of Management Sciences, Bahria University, Islamabad
Sakina_ali835@hotmail.com

Dr. Muhamamd Anees Khan
Associate Professor/ HoD Department of Management Sciences, Bahria University, Islamabad
manees.buic@bahria.edu.pk

Abstract

Stock market prediction is crucial for capital allocation and economic growth, but it is also challenging due to the uncertainty and complexity of the future. This study compares the performance of four machine learning models - Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Recurrent Neural Network (RNN), and Long-short term memory (LSTM) - in predicting the stock price in different geographical locations (S&P500, NYSE, NASDAQ, SSEC, EURONEXT, TSE). Using daily stock price data from November 25th, 2012, to November 25th, 2022, the models are evaluated based on root mean square error (RMSE), mean bias error (MBE), accuracy, and mean absolute percentage error (MAPE) and cross-validating the prediction results. The results show that KNN consistently outperforms the other models in most regions with the lowest error rates and the highest accuracy. The cross-validation results further confirm the superiority of KNN over the other three models.

Keywords: *Artificial intelligence, Artificial Neural Network, K-Nearest Neighbor, Recurrent Neural Network, and Long-short term memory, Stock Prediction, RMSE, MBE*

1. Introduction

The stock market, with its 400-year history of trading, transferring, and circulating funds, plays a pivotal role in the global economy. Organizations depend heavily on the capital generated by these markets, thereby underscoring the importance of market analysis(J. Zhang et al., 2021). The influx of substantial capital into the stock market bolsters the organic composition of corporate capital, thereby fostering economic development(Wu et al., 2022). To increase the flow of capital, investors need to allocate significant sums of funds to stimulate economic growth. However, they must also ensure the protection of their investment given the volatile character of the stock market. Therefore, stock market prediction has attracted significant attention from both industry and academia, as it serves as an indicator of a country's economic activities. The level of accuracy in predictions has a direct impact on the profitability of investment strategies, portfolio management, and trading activities. It allows investors to detect prospective possibilities, reduce risks, and maximize their investment decisions(Mehtab et al., 2021). Literature abounds with traditional and machine learning models employed for stock market prediction. Each model strives to precisely forecast stock values, aiding investors in identifying prospective investment opportunities(Vijh et al., 2020). Machine learning models outperform traditional models in predicting stock values due to their superior ability to process and analyze huge and complicated datasets. They can capture complex patterns, recognize non-linear correlations, and extract valuable characteristics from the data that may not be apparent using traditional statistical methods. Machine learning models, such as neural networks, decision trees, or support vector machines, are specifically built to handle non-linearities and can capture complex patterns and correlations in the data. Studies have demonstrated that machine learning (ML) models can accurately predict stock prices with an accuracy range of 60-86%. (Li et al., 2017; Vijh et al., 2020).

There are many machine learning models used for prediction purposes. Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Networks (ANN), Multilayer Perception (MPL), Neural Networks (RNN), Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), etc., are some of the powerful models for processing sequential data that help analyze the hidden patterns and underlying dynamics to predict stock prices. With the rapidly growing volume of data and computational resources, machine learning models have provided fantastic performances. However, no studies have resulted in the use of any specific model that can be used for prediction purposes. (Li et al., 2017; S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, Vijay Krishna Menon, and K. P. Soman, 2017; Vijh et al., 2020). In light of the significant importance of stock price prediction and the magnificent accuracy exhibited by machine learning models, this research aims to compare the performances of four mostly used neural models (ANN, KNN, RNN, and LSTM) to determine the most effective machine learning model for predicting stock prices in six different regions: S&P 500,National Association of Securities Dealers Automated Quotations(NASDAQ), Shanghai Composite (SSEC), New York Stock Exchange (NYSE), Euronext, and Tokyo Stock Exchange (TSE). This may assist investors in identifying the most efficient model that can empower investors or institutions to make more informed investment decisions and achieve favorable outcomes (Pawaskar, 2022; Walayat Hussain, 2022).

The results of this study have important consequences for policymakers, financial institutions, traders, investors, financiers, and anyone with a vested interest in investments. By using the results generated by these models and conducting performance evaluations, individuals can choose the most precise predictive model for their desired investment area, thereby achieving higher returns. By utilizing the most efficient machine learning model, policymakers can reduce the risk of highly volatile stock markets and develop methods that yield significant financial returns for all investors. To summarize, this paper makes an academic contribution by highlighting the importance of machine learning models. This may enable researchers to shift their focus towards more advanced prediction approaches rather than conventional methods. Furthermore, it will enhance the existing academic understanding of non-linear financial time series.

2. LITERATURE REVIEW

Literature is full of such studies where historical prices and fundamental values have been used to predict the future prices using machine learning models(Lo & MacKinlay, 1988). Some of such studies are as follows.

2.1.Artificial Neural Network

ANN is a computer structure that functions similarly to biological neurons in terms of performance(Moghaddam et al., 2016). It is considered a non-linear statistical data tool because they can stimulate the link between inputs and outputs (Rather et al., 2015). The primary benefit of ANN is its capacity to infer fundamental patterns from the data, where most traditional approaches fall short(G. Zhang et al., 1998). A typical ANN consists of three layers each serving a specific purpose in the learning and prediction process. The most common types of layers in an ANN model are:

1. **Input layer:**The input layer is the initial layer of the neural network, which receives the input data. Each node in the input layer represents a feature or attribute of the input data(Gurjar et al., 2018).
2. **Hidden layer:**Hidden layers are intermediate layers between the input and output layers. They play a crucial role in learning complex patterns and extracting higher-level representations from the input data. A neural network may have one or more hidden layers, depending on the complexity of the problem and the desired model architecture(Chhajer et al., 2022).
3. **Output Layer:** The output layer is the final layer of the neural network, responsible for producing the predicted output or class labels. The number of nodes in the output layer depends on the nature of the prediction task(Aslam et al., 2022).Except for the input layer, all nodes in the hidden and output layers use non-linear activation functions. Each node in the input layer is connected to every neuron in the hidden layer, which is followed by the output layer (M et al., 2018).
4. **Activation Functions:** Activation functions are applied to the nodes in the hidden and output layers to introduce non-linearity into the model. Non-linear activation functions allow neural networks to learn and represent complex relationships between the input and output. Examples of popular activation functions include the sigmoid function, ReLU (Rectified Linear Unit), and softmax function(Gurjar et al., 2018; Ravichandran et al., 2007).

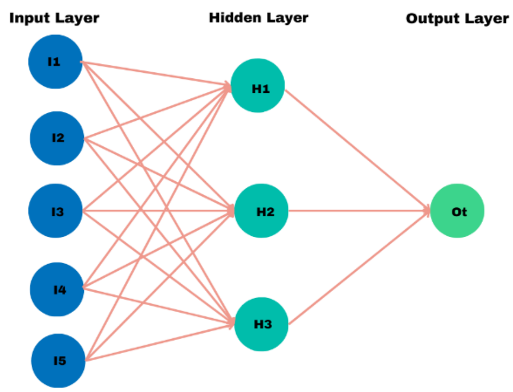

Figure No. 1. Artificial Neural Network

Numerous research examining the movement patterns of various financial instruments have been conducted in recent years(Kara et al., 2011). Artificial neural networks, or ANNs, are algorithms created to comprehend complex problems that are beyond the capabilities of accessible neural networks or simple ML algorithms (Dhenuvakonda et al., 2020). It is one of the data mining techniques that is gaining popularity due to its ability to recognize and understand the nonlinear variable’s relationship(X. Wu et al., 2001). As per Avci (2007) the first notable research on the use of neural network models for stock price prediction was of White (1988). After White's study, several investigations of neural network models were conducted. Kara et al. (2011) used two models based on the ANN and SVM to discuss the movement of the stock prices. They evaluated the performances of the two models and found that, on average, the ANN model performed much better than the judgments made by the SVM model. As perthe experimental findings, the average performance of the ANN model (75.74%) was shown to be substantially higher than the SVM model (71.52%). Similar study was later conducted by Vijh et al. (2020) while comparing the accuracy level of ANN and Random forest. The comparative research based on MAPE, RMSE and MBE values confirms the better performance of ANN over Random Forest.

2.2.K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN), a supervised, non-parametric controlled learning method which uses proximity, to classify or predict the arrangement of a set of individual data points. It can be used to solve classification as well as regression problems. After determining each distance, the algorithms label the data based on the nearest neighbors (Yunneng, 2020). The stock market, with its extensive and dynamic information sources, presents an ideal environment for data mining and business research. Previously, the K-Nearest Neighbor (KNN) algorithm and a non-linear regression approach were employed to predict stock prices for a sample of six major Jordanian companies listed on the Jordanian stock exchange. This was done to aid investors, management, decision-makers, and users in making informed investment decisions. The KNN method proved robust, with a low error ratio, leading to rational and reasonable findings. Moreover, the forecast results, based on real stock price data, closely paralleled the actual stock prices (Alkhatib et al., 2013).

In a similar vein, a hybridized Support Vector Machine (SVM) and K-Nearest Neighbor framework was used to forecast the Indian stock market. The goal was to gain a comprehensive understanding of the market in the Indian context, focusing on two indices, the Bombay Stock Exchange (BSE Sensex) and CNX Nifty. This was achieved using technical analysis techniques and tools to predict the closing price, volatility, and momentum of the stock market based on available data. The hybrid model predicted profit or loss using SVM with various kernel functions. The output of SVM also assisted in computing the best closest neighbor from the training set to predict future stock values over time horizons of one day, one week, and one month. The proposed SVM-KNN hybrid model scaled well to high-dimensional data, and the trade-off between classifier complexity resulted in better prediction capability (Nayak et al., 2015).

Subsequent studies reviewed supervised machine learning models to discuss how these models are applied to enhance the accuracy of stock price prediction. Due to its superior performance and accuracy, the Support Vector Machine (SVM) was found to be the most commonly used approach for stock price prediction. Positive prediction results were also demonstrated by other approaches such as the Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Naive Bayes, Random Forest, Linear Regression, and Support Vector Regression (SVR) (Lawal et al., 2020). The KNN technique, capable of understanding relationships between numeric data, proved powerful in numeric prediction issues for anticipating next-day changes in stock value. Studies have shown a superior accuracy of 70% when using these algorithms in stock market prediction (Latha et al., 2022).

2.3.Recurrent Neural Network/LSTM

"Recurrent Neural Networks (RNNs) are a class of algorithms that have gained prominence due to their unique internal memory feature. Although they were initially developed in the 1980s, their usage was not widespread until the 1990s, when the introduction of Long Short-Term Memory (LSTM) and the availability of large datasets brought RNNs to the forefront of machine learning research. The key advantage of RNNs is their ability to remember crucial aspects of the input, enabling them to make precise predictions about future contexts. In this type of neural network, all inputs and outputs are independent of each other. However, when necessary, the output from a previous step (stored in the memory) is used as an input for the current step. This characteristic has led to the wide use of RNNs in prediction tasks(Kumar et al., 2018; Saud & Shakya, 2020).

RNNs are deep learning models designed to analyze sequential data. Unlike feedforward neural networks, RNNs have feedback connections that allow them to retain information from past inputs. This makes them ideally suited for tasks such as natural language processing, speech recognition, and time series prediction1. The operation of RNNs is based on preserving the output of a certain layer and feeding it back to the input to predict the output of that layer (Pawar et al., 2019). RNNs are a type of neural network in which the units are recurrently linked. This enables them to process a series of inputs using their internal memory. As a result, they can be used for tasks such as handwriting recognition, text production, stock market prediction, and speech recognition(Dewan & Sharma, 2015; Sak et al., 2014).

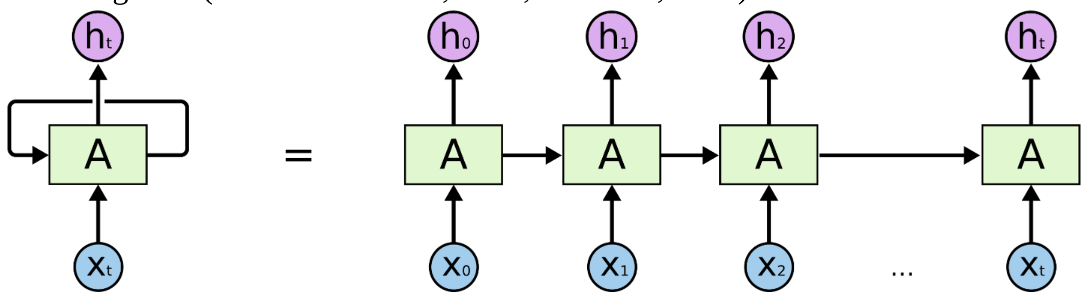

Figure No. 2. Illustration of RNN

LSTMs are a type of RNN architecture specifically designed to handle long-term dependencies more accurately than traditional RNNs. They have proven to be effective in tasks such as forecasting changes in stock data and voice recognition, outperforming Deep Neural Networks (DNNs) and basic RNN models(Sak et al., 2014).One of the key features of LSTM is the introduction of memory cells in the network's hidden layer. These computational units, which replace conventional artificial neurons, allow the network to efficiently connect memories and distant inputs. This makes them adept at understanding the evolving structure of data over time, thereby enhancing their predictive capabilities.In the realm of sequential data processing, RNNs have demonstrated their effectiveness. The performance of LSTM and RNN was evaluated using five years of NIFTY 50 stock data, with the model's Root Mean Square Error (RMSE) values indicating that both RNN and LSTM provide reliable forecasting (Roondiwala et al., 2017).An LSTM model with seven hidden layers was compared with Support Vector Regression (SVR) using data from January 2015 to January 2020. The Mean Absolute Percentage Error (MAPE) of the LSTM and SVR was compared across several stock indices, including the NSE, BSE, NASDAQ, NYSE, S&P 500, and Dow Jones Industrial Average. The results showed that LSTM outperformed SVR, providing superior prediction accuracy(Bathla, 2020).

Similarly, an RNN-LSTM model was used in a study on NIFTY-50 stocks, considering four characteristics (high, close, open, and low price of each day). Five years of data was used for prediction, with RMSE as the error measure to be minimized via backpropagation. The LSTM model's hyperparameters were tuned using the grid-search method to ensure that validation losses stabilize with an increasing number of epochs and that the validation accuracy converges. The study employed four alternative models, each with a different architecture and input data format, to leverage the predictive potential of LSTM regression models in predicting future NIFTY 50 open values. The results clearly showed that the most accurate model was the LSTM-based univariate model, which uses one-week past data as input to forecast the open value of the NIFTY 50 time series for the next week (Mehtab et al., 2021).

3. RESEARCH DATA AND METHODOLOGY

This study utilizes the daily historical prices of the six leading stock exchanges worldwide: the New York Stock Exchange (NYSE) with a market capitalization of $25.23 trillion, NASDAQ ($20.58 trillion), Shanghai Stock Exchange ($6.6 trillion), Euronext ($6.26 trillion), S&P 500, and the Tokyo Stock Exchange ($5.75 trillion)(Henrique et al., 2018; Kumar et al., 2018) (Henrique et al., 2018; Kumar et al., 2016). The data, sourced from investing.com, covers a ten-year period from November 24, 2012, to November 24, 2022. It includes key information about the stock prices, such as the High, Low, Open, and Closing prices of the indexes.The study employs a specific data analysis methodology, which is illustrated in the subsequent flow chart. Four Machine Learning (ML) models - Artificial Neural Network (ANN), K-Nearest Neighbors (KNN), Recurrent Neural Network (RNN) and Long-Short term Memory (LSTM) are used to process the data and evaluate their accuracy levels.

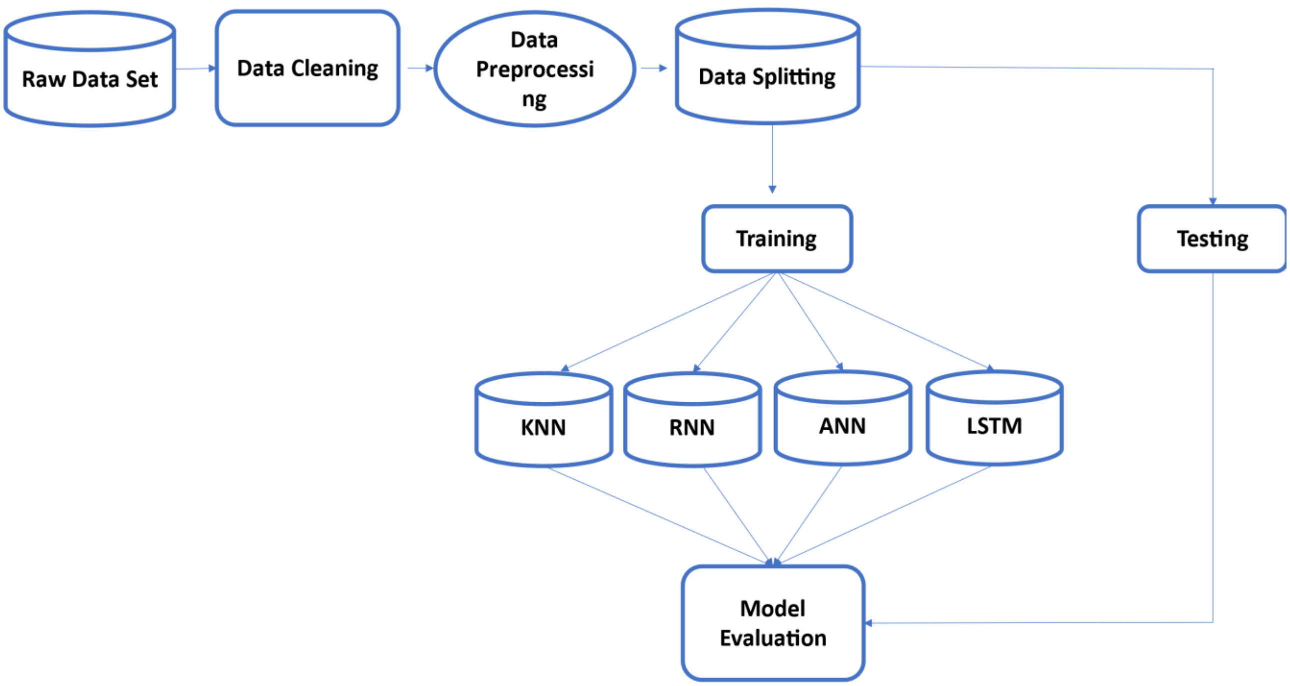

Figure No. 3. Machine Learning Process

The process of machine learning evaluation starts with cleaning the imported data, which involves identifying and handling missing and duplicate values. The primary goal of this step is to minimize errors that could impact the final analysis and the results. After the data is cleaned, it is processed for descriptive statistics. The descriptive statistics of the data used in the study are explained in Table No. 1 below.

Table No. 1. Descriptive Statistics

	Open		High		Low		price		Count
	Mean	**std**	**mean**	**Std**	**mean**	**std**	**Mean**	**std**	
NYSE	12212.27	2201.61	12272.23	2218.98	12147.58	2183.59	12214.07	2201.18	2517
S&P500	2626.97	862.17	2640.39	867.70	2612.33	855.84	2627.38	862.01	2453

NASDAQ	7546.49	3513.82	7596.26	3544.12	7490.73	3476.90	7547.03	3511.98	2518
SSEC	3002.70	535.60	3025.96	541.90	2980.19	526.85	3006.15	535.72	2432
TSE	1774.16	226.37	1784.36	226.19	1763.58	226.49	1774.43	225.96	2442
EURONEXT	990.77	165.34	996.25	166.00	984.85	164.83	990.80	165.43	2560

Notes: NYSE = New York Stock Exchange, NASDAQ= National Association of Securities Dealers Automated Quotations, SSEC= Shanghai Stock Exchange Composite Index, TSE= Tokyo Stock Exchange

The following table provides an explanation of the key data features of the top six global indices: open, low, high, and closing prices. As ML algorithms operate more effectively with comparatively smaller and closer to normal distribution features, therefore the data features must be preprocessed and translated to a given range between 1 and 0. Due to the lengthier training times and lower accuracy levels of data sets with huge discrepancies, this enables the features' values to train correctly and fast, producing highly accurate results. Consequently, in order to achieve the utmost precision, scaling the data point is a necessary step because it is a generalized strategy used to bring the data closer together. Following the features of the data used in the study, MinMaxScaler is used to normalize the data and scales the data to a specific scale between and. 0 and 1.

In machine learning and statistical modeling, the train-test split refers to the process of dividing a dataset into two separate subsets: the training set and the testing set (also known as the validation set or the holdout set).The training set is used to train or adapt the model, enabling it to discover patterns, relationships, and parameters from the data. It is the portion of the dataset used by the model to adjust its internal parameters and generate predictions.In contrast, the testing set is utilized to evaluate the performance of the trained model. It functions as a sample of data that the model was not exposed to during its training. We can estimate how well the model is likely to generalize to new, unseen data by evaluating its performance on the testing set.

Typically, the train-test split is performed by randomly dividing the dataset into two portions while preserving the overall distribution of the data. Commonly, a larger portion of the data is allocated to the training set (e.g., 70-80%) and a lesser portion to the testing set (20-30%). The study uses the commonly splitting techniques of 70:30 where 70% of the data is used to train while the other 30 is used to test the model’s accuracy.

4. MODEL APPLICATION AND EVALUATION

As per the process explain in Figure No.3, after the data is splitin to train and test as per the guidelines, all the machine learning models are applied. The training set is used to train the all the models (ANN, KNN, RNN, LSTM) while the test set is used to assess the final performance of the trained model. As per the process the four models of the study are applied individually across all the regions (NYSE, NASDAQ, S&P 500, SSEC, TSE, and EURONEXT) and each model is tested across all the regions to identify the best model for the specific region. As the models are trained and performances are tested, each model must be evaluated for the error values to check their accuracy rates.

4.1 Model Evaluation

Evaluation of machine learning models is essential for assessing their performance and determining their efficacy in completing a particular task. It is a critical step in machine learning to assess the performance and quality of a trained model. It helps you understand how well the model generalizes to unseen data and whether it meets the desired requirements of your problem. In machine learning, the following evaluation metrics and methodologies are prevalent:

4.1.1 Root Mean Square Error

Root Mean Square Error (RMSE) is a popular metric in machine learning for assessing regression models. It calculates the square root of the average squared differences between predicted and actual values, providing a measure of the typical prediction error magnitude(Latha et al., 2022).It can be expressed as,

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n}(yi - y\hat{}i)2}}{n} \text{………………………….. } \textit{equation no.1}$$

In the above equation, n is the number of data points, y(i) is the i-th measurement, and y ̃(i) is its corresponding prediction.

4.1.2 Mean Bias Error

Mean Bias Error (MBE), also referred to as Mean Error (ME) or Mean Residual Error, is a metric used in regression analysis. It quantifies the average discrepancy between the predicted and actual values. MBE offers insights into the overall bias or systematic error present in the model's predictions(Chhajer et al., 2022). it can be expresses as

$$MBE = 1/n\sum_{i=1}^{n}(yi - \mathrm{y\hat{}i}) \text{……………………} \textit{equation no.2}$$

yi represents the actual values in the data set, y ̂i represents the predicted values and N is the number of data points.

4.1.3 Mean Absolute percentage error

It is a statistical approach used to determine the precision of a machine learning model through the absolute proportion of predicted errors which is defined as the difference between the actual and predicted values (Patel & Wang, 2015). It is a widely used indicator to assess a forecasting model's accuracy. MAPE measures the average absolute percentage difference between the actual and projected values. It can statistically be expressed as,

$$\mathrm{MAPE}{=}1/\mathrm{n}\sum_{i=1}^{n}(yi - \mathrm{yi\hat{}})/\mathrm{yi} \text{……………………} \textit{equation no.3}$$

yi represents the actual values, y ̂i represents the predicted values and N is the number of observations.

4.1.4 Cross Validation

Cross-validation is a statistical method used to evaluate and compare the performance of machine learning algorithms. It involves partitioning the data into two segments: a training set for model development, and a validation set for model assessment (Lei, 2017). The main goal of cross-validation is to prevent overfitting, a situation where a model, having been overly trained on the training data, performs poorly on new, unseen data.

By evaluating the model on multiple validation sets, cross-validation improves the accuracy of the model's generalization performance estimate. This assessment gauges the model's ability to perform effectively on unfamiliar data. K-fold cross-validation is the most basic form of cross-validation that involves multiple rounds of k-fold cross-validation (Stone, 1974). In the k-fold cross-validation method, the dataset is randomly split into a series of equal-sized subsets or folds. The number of these subsets is represented by the value of k. In this study, k is set to 5, meaning the dataset is divided into 5 separate subsets. Four of these subsets are used for training the model, while the remaining subset is used for testing. This process is repeated 5 times, with each

iteration using a different subset as the test set and the remaining subsets as the training data. The average of the results from these iterations is then calculated (Refaeilzadeh et al., 2009).
The results of all the evaluation techniques are presented in the table below.

Table No.2. Performance Evaluation

	KNN				ANN				RNN		
Region	**RMSE**	**MBE**	**MAPE**	**Cross Validation**	**RMSE**	**MBE**	**MAPE**	**Cross Validation**	**RMSE**	**MBE**	**MAPE**
S&P 500	0.005	0.008	0.014	97.90%	0.02	0.003	0.084	96.5%	0.011	0.002	0.041
NYSE	0.007	0.002	0.011	96.40%	0.015	0.002	0.037	92.10%	0.011	0.001	0.753
NASDAQ	0.007	0.008	0.016	93.20%	0.012	0.001	0.03	90.30%	0.011	0.007	0.554
SSEC	0.008	0.003	0.024	98.40%	0.012	0.001	0.186	87.40%	0.008	0.01	0.299
TSE	0.009	0.002	0.013	97.90%	0.023	0.001	0.031	98.30%	0.013	0	0.569
EURONEXT	0.011	0.004	0.014	96.35%	0.013	0.001	0.037	94.10%	0.489	0.439	0.569

Notes: NYSE = New York Stock Exchange, NASDAQ= National Association of Securities Dealers Automated Quotations, SSEC= Shanghai Stock Exchange Composite Index, TSE= Tokyo Stock Exchange

	LSTM			
Region	**RMSE**	**MBE**	**MAPE**	**Cross-Validation**
S&P 500	0.244	0.201	0.031	91.90%
NYSE	0.008	0.001	0.041	94.40%
NASDAQ	0.503	0.391	0.032	90.20%
SSEC	0.015	0.001	0.0915	92.40%
TSE	0.011	0.006	0.962	78.90%
EURONEXT	0.009	0.005	0.408	81.35%

While comparing the models for predicting S&P 500 performance, KNN outperforms with the lowest RMSE, MAPE, and MBE values (0.005, 0.041, and 0.008 respectively) and the highest cross-validation score of 97.90%. RNN, however, yields the least favorable results.It’s important to remember that stock price predictions, especially for the S&P 500, carry inherent uncertainties and accuracy isn’t guaranteed. Investors should consider forecasts as part of a holistic strategy that includes fundamental analysis, diversification, and risk management. Consulting financial experts can provide valuable insights for evaluating projections and making investment decisions. Similar results are seen when comparing models for predicting NYSE, NASDAQ, SSECand EURONEXT performance, KNN emerges as the superior model with the lowest RMSE, MAPE, and MBE values and the highest cross-validation score of 96.4%, 93.20%, 98.40% and 96.35%. Conversely, RNN delivers the least impressive results. However, ANN has the best results while predicting the stock prices of TSE with highest error rates of prediction by LSTM model.

The tabular representation additionally validates a slight discrepancy between the predicted and actual stock price values generated by the KNN model for S&P500, NYSE, SSEC, and EURONEXT as explained in the table below.

Table No.3 Actual verses Predicted Prices

	S&P 500					NYSE		
Date	**Actual Price**	**ANN Predicted Price**	**KNN Predicted Price**	**RNN Predicted Price**	**LSTM Predicted Price**	**Actual Price**	**ANN Predicted Price**	**KNN Predicted Price**
11/25/2022	4132.15	4141.54	4126	4136.1	4141.1	15545.5	15396.73	15112.19
11/24/2022	4158.24	4139.24	4145.2	4144.84	4149.84	15481.8	15233.54	14868.65
11/23/2022	4057.84	4053.79	4044.7	4038.74	4043.74	15278.3	15165.23	14673.8
11/22/2022	3978.73	3980.5	3970.3	3966.26	3961.26	15309.8	15182.99	14718.67
11/21/2022	3941.48	3919.74	3925.02	3932.75	3927.75	15224	15001.06	14500.79

	SSEC					EURON		
Date	**Actual Price**	**ANN Predicted Price**	**KNN Predicted Price**	**RNN Predicted Price**	**LSTM Predicted Price**	**Actual Price**	**ANN Predicted Price**	**KNN Predicted Price**
11/25/2022	3101.09	3090.9	3098.2	3085.2	3075.2	1279.52	1274.91	1277.52
11/24/2022	3089.31	3095.58	3085.3	3065.3	3055.3	1278.56	1272.92	1276.56
11/23/2022	3096.91	3088.88	3089.63	3075.6	3065.6	1273.2	1266.62	1271.2
11/22/2022	3088.94	3095.36	3084.5	3065.2	3055.2	1266.02	1257.76	1264.02
11/21/2022	3085.04	3073.87	3081.5	3076.65	3066.65	1256.95	1257.09	1254.95

	NASDAQ					TSE		
Date	**Actual Price**	**ANN Predicted Price**	**KNN Predicted Price**	**RNN Predicted Price**	**LSTM Predicted Price**	**Actual Price**	**ANN Predicted Price**	**KNN Predicted Price**
11/25/2022	12081.39	12046.44	12075.3	1207.1	1205.25	1965.06	1966.41	1878.84
11/24/2022	12131.13	11847.25	12125.32	12124.32	12111.36	1965.73	1957.92	1872.82
11/23/2022	11740.65	11441.16	11425.35	11726.32	11702.56	1953.75	1970.71	1874.53
11/22/2022	11434.74	11232.75	11422.3	11426.7	11414.68	1967.05	1959.86	1877.57
11/21/2022	11264.45	11256.39	11256.35	11258.2	11225.2	1960.25	1958.56	1945.54

ANN: Artificial Neural Network, KNN: K-Nearest Neighbor, RNN: Recurrent Neural Network, LSTM: Long-Short Term memory

CONCLUSION

Modern stock markets significantly impact the wider economy, as businesses heavily depend on the revenue they generate, underscoring the importance of market analysis. However, predicting stock price movements is complex due to the interplay of various factors resulting into constant fluctuations in stock prices which increases investment risk and economic losses, ultimately affecting the country’s economic structure. Therefore, accurate predictions of stock price can offer investors numerous opportunities to boost their investment returns. Machine learning models have proven to be powerful tools, with previous studies highlighting their exceptional predictive abilities and high accuracy levels.

This paper primarily aimed to use the Artificial Neural Network (ANN), K-nearest Neighbor (KNN), Recurrent Neural Network (RNN) and Long-short Term Memory (LSTM) to predict the daily prices of six different global indices. The machine learning model was used to predict the stock prices of the Shanghai Composite (SSEC), New York Stock Exchange (NYSE Composite), Euronext 100 (N100), Tokyo Stock Exchange (TSE), S&P 500 (SPX), and NASDAQ Composite (IXIC) using the daily high, low, and open prices from November 25th, 2012, to November 25th, 2022. The models were evaluated using widely accepted performance indicators, namely root mean square error (RMSE), mean bias error (MBE), accuracy, and mean absolute percentage error (MAPE). The results showed that the K-nearest Neighbor (KNN) models consistently outperformed the other three models in the majority of the analyzed regions with the lowest error rates of RMSE, MAPE, and MBE. The cross-validation results further confirmed the higher accuracy rates of KNN, followed by ANN, in predicting stock prices across all six regions, thus demonstrating the superiority of these two models. These findings could help high-frequency traders improve their ability to predict stock returns and covariances in these markets. Future research efforts may focus on exploring other machine learning models to conduct a comparative analysis of their respective accuracy levels. The goal of such studies would be to identify the most optimal machine learning model for predicting the prices of widely traded stock markets worldwide.

REFERENCES

Alkhatib, K., Najadat, H., Hmeidi, I., &Shatnawi, M. (2013). *Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm*. https://www.semanticscholar.org/paper/Stock-Price-Prediction-Using-K-Nearest-Neighbor-Alkhatib-Najadat/1507329f5382a1550430657ebae7f70507f63410

Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. *Research in International Business and Finance*, *62*, 101744. https://doi.org/10.1016/j.ribaf.2022.101744

Avci, E. (2007). FORECASTING DAILY AND SESSIONAL RETURNS OF THE ISE-100 INDEX WITH NEURAL NETWORK MODELS. *Doğuş ÜniversitesiDergisi*, *8*(2), Article 2.

Bathla, G. (2020). Stock Price prediction using LSTM and SVR. *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 211–214.

Chhajer, P., Shah, M., & Kshirsagar, A. (2022). The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction. *Decision Analytics Journal*, *2*, 100015. https://doi.org/10.1016/j.dajour.2021.100015

Dewan, A., & Sharma, M. (2015). Prediction of heart disease using a hybrid technique in data mining classification. *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, 704–706.

Dhenuvakonda, P., Anandan, R., & Kumar, N. (2020). Stock price prediction using artificial neural networks. *Journal of Critical Reviews*, *7*(11), 846–850.

Gurjar, M., Naik, P., Mujumdar, G., & Vaidya, T. (2018). Stock market prediction using ANN. *International Research Journal of Engineering and Technology*, *5*(3), 2758–2761.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*, *4*(3), 183–201. https://doi.org/10.1016/j.jfds.2018.04.003

Kara, Y., Acar Boyacioglu, M., &Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, *38*(5), 5311–5319. https://doi.org/10.1016/j.eswa.2010.10.027

Kumar, I., Dogra, K., Utreja, C., & Yadav, P. (2018). A comparative study of supervised machine learning algorithms for stock market trend prediction. *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 1003–1007.

Latha, R. S., Sreekanth, G. R., Suganthe, R. C., Geetha, M., Selvaraj, R. E., Balaji, S., Harini, K. R., & Ponnusamy, P. P. (2022). Stock Movement Prediction using KNN Machine Learning Algorithm. *2022 International Conference on Computer Communication and Informatics (ICCCI)*, 1–5. https://doi.org/10.1109/ICCCI54379.2022.9740781

Lei, J. (2017). *Cross-Validation with Confidence* (arXiv:1703.07904). arXiv. http://arxiv.org/abs/1703.07904

M, H., E.a., G., Menon, V. K., &K.p., S. (2018). NSE Stock Market Prediction Using Deep-Learning Models. *Procedia Computer Science*, *132*, 1351–1362. https://doi.org/10.1016/j.procs.2018.05.050

Mehtab, S., Sen, J., & Dutta, A. (2021). Stock price prediction using machine learning and LSTM-based deep learning models. *Symposium on Machine Learning and Metaheuristics Algorithms, and Applications*, 88–106.

Moghaddam, A. H., Moghaddam, M. H., &Esfandyari, M. (2016). Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science*, *21*(41), 89–93. https://doi.org/10.1016/j.jefas.2016.07.002

Nayak, J., Naik, B., & Behera, Prof. Dr. H. (2015). *A comprehensive survey on support vector machine in data mining tasks: Applications & challenges*. *8*, 169–186. https://doi.org/10.14257/ijdta.2015.8.1.18

Patel, N., & Wang, J. T. L. (2015). Semi-supervised prediction of gene regulatory networks using machine learning algorithms. *Journal of Biosciences*, *40*(4), 731–740. https://doi.org/10.1007/s12038-015-9558-9

Ravichandran, K. S., P.Thirunavukarasu, R, N., &R.Babu. (2007). Estimation of return on investment in share market through ANN. *Journal of Theoretical and Applied Information Technology*, *3*.

Roondiwala, M., Patel, H., & Varma, S. (2017). Predicting Stock Prices Using LSTM. *International Journal of Science and Research (IJSR)*, *6*. https://doi.org/10.21275/ART20172755

Sak, H., Senior, A. W., &Beaufays, F. (2014). *Long short-term memory recurrent neural network architectures for large scale acoustic modeling*.

Saud, A. S., & Shakya, S. (2020). Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE. *Procedia Computer Science*, *167*, 788–798.

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Computer Science*, *167*, 599–606. https://doi.org/10.1016/j.procs.2020.03.326

White, H. (1988). Economic prediction using neural networks: The case of IBM daily stock returns. *ICNN*, *2*, 451–458.

Wu, S., Liu, Y., Zou, Z., & Weng, T.-H. (2022). S_I_LSTM: Stock price prediction based on multiple data sources and sentiment analysis. *Connection Science*, *34*(1), 44–62. https://doi.org/10.1080/09540091.2021.1940101

Wu, X., Fund, M., &Flitman, A. (2001). Forecasting stock performance using intelligent hybrid systems. *Springerlink*, *44*(7), 447–456.

Yunneng, Q. (2020). A new stock price prediction model based on improved KNN. *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*, 77–80. https://doi.org/10.1109/ICISCE50968.2020.00026

Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

Zhang, J., Li, L., & Chen, W. (2021). Predicting Stock Price Using Two-Stage Machine Learning Techniques. *Computational Economics*, *57*(4), 1237–1261. https://doi.org/10.1007/s10614-020-10013-5