DOI: https://doi.org/10.52015/nijbm.v19i1.197

NUML International Journal of Business & Management
Volume 19, Issue 1, June (2024)
Journal Home Page: nijbm.numl.edu.pk/index.php/BM
ISSN 2410-5392 (Print), ISSN 2521-473X (Online)

Stock Price Prediction Using Machine Learning: Evidence from Pakistan Stock Exchange

Zafar Akhter[1], Hassan Raza[2]*,

Article History:
Received: February 24,2024
Revised: April 14, 2024
Accepted: June 1, 2024
Available Online: June 30, 2024

Keywords:
Machine learning; Stock prices; Random Forest; Precision; Recall, Financial Forecasting.

Funding:
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Abstract

This research investigates the utilization of machine learning methodologies for the purpose of forecasting the fluctuations in stock values inside the financial market. The application of a Random Forest classifier is utilized on a dataset including historical stock prices (namely, the KSE-100 Index) to generate predictions regarding the future movement of stocks, specifically whether they would experience an increase or decrease. The model is trained via a sliding window methodology and is assessed through the utilization of precision, recall, and F1-score criteria. The study furthermore incorporates the utilization of back testing and hyper-parameter tweaking techniques in order to enhance the performance of the model. The findings indicate that the model demonstrates a precision score of 58%, representing an enhancement compared to the previous score of 48%. Nevertheless, the model's total accuracy stands at a mere 58%, underscoring the need for future enhancements. The report additionally proposes potential avenues for future research, such as exploring alternate data sources, employing sentiment analysis techniques, and developing more advanced algorithms. The findings of this study hold significant significance for investors and financial institutions, as they highlight the potential of machine learning in facilitating informed investment decisions and improving financial forecasts and analysis.

INTRODUCTION

The stock market is a complex and volatile realm that is influenced by a multifaceted array of elements, such as economic indicators, company performance measurements, political events, and investor sentiment. The intricate nature of this situation highlights the importance of making precise forecasts regarding changes in stock prices. This effort is crucial for investors who aim to optimize their gains while minimizing potential losses. Past studies indicate that investors usually predominantly utilized traditional forecasting methodologies like as basic and technical analysis. Nevertheless, the precision and

[1] PhD Scholar, Management Sciences, SZABIST University, Islamabad, Pakistan
* Corresponding Author: dr.hassan@szabist-isb.edu.pk
[2] Associate Professor, Management Sciences, SZABIST University, Islamabad, Pakistan

dependability of these conventional approaches are constrained, frequently falling short in capturing the complete extent of market volatility.

The utilization of machine learning algorithms in the prediction of stock prices signifies a notable progression in the field of financial analytics, presenting the possibility of enhanced accuracy in forecasts and economic advantages for investors. These algorithms analyze large historical datasets, which include market indicators, fundamental analysis, sentiment evaluations, and risk assessments, in order to detect patterns and correlations that may not be readily apparent using conventional analysis techniques (Gursida, 2017; Mehtab & Sen, 2020, Aziza et al., 2021; Chen, 2023; Nafia et al., 2023; Safari & Badamchizadeh, 2023). In contrast to traditional regression methods, machine learning has exceptional proficiency in identifying intricate, non-linear relationships among predictors, frequently resulting in higher efficacy in predicting stock values. Furthermore, these algorithms demonstrate exceptional proficiency in identifying crucial prognostic indications, including momentum, liquidity, and volatility. This significantly contributes to our comprehension of asset pricing and facilitates the advancement of financial innovation. They possess the capability to process extensive, unorganized datasets, so offering a more profound understanding of market dynamics and the complex interaction of multiple components.

The major gains in stock price prediction facilitated by machine learning, such as enhanced accuracy and the identification of crucial predictive signals, necessitate a heightened awareness of their inherent limitations (Han & Fu, 2023; Ruan et al., 2020).

In recent years, these significant breakthroughs have played a pivotal role in driving the progress of novel forecasting approaches. These innovative methodologies, utilizing state-of-the-art technology and advanced algorithms, provide greatly improved precision, therefore offering the potential for considerably greater profits for individuals involved in the market. The advancement in market analysis signifies a significant progression in the ability to predict financial outcomes, while also ushering in a fresh era of investment tactics that are guided by data and analytics.

This study explores the emerging domain of utilizing machine learning algorithms for the purpose of analyzing financial markets, specifically emphasizing their effectiveness in predicting stock values. The focal point of our study revolves around the development of an advanced machine learning model with the purpose of effectively predicting the directional movement of stock prices. More specifically, our aim is to estimate whether these prices will experience an upward or downward trend in the following trading session. The primary aim of our research is to improve the precision of the model in predicting true positives, which refers to cases where the algorithm accurately forecasts an increase in a stock's price and this increase actually occurs.

Access to vast quantities of high-caliber data is necessary for machine learning models, yet they are susceptible to dangers such as overfitting. The continuous examination and tweaking of models are vital in order to uphold their efficacy. While machine learning has the potential to provide improvements compared to conventional linear regression models in certain situations, its effectiveness can vary and its suitability may not always be dependent on the specific characteristics of the problem being addressed. However, the incorporation of machine learning techniques in the prediction of stock prices has the potential to revolutionize financial forecasting. This highlights the importance of striking a delicate equilibrium between adopting innovative approaches and recognizing the inherent difficulties associated with these sophisticated analytical instruments.

In light of this, the following research questions and corresponding hypotheses have been formulated to contribute to the existing body of research literature.

1. Can machine learning algorithms enhance the precision of stock price predictions in comparison to conventional forecasting methods?

Hypothesis 1: Machine learning algorithms will yield more precise stock price predictions compared to conventional approaches.

2. Can machine learning algorithms accurately detect non-linear correlations and patterns in stock market data?

Hypothesis 2: Machine learning models have the ability to detect non-linear relationships and patterns in stock market data, resulting in enhanced prediction performance.

Theoretical Framework

The present work aims to investigate the fundamental theoretical underpinnings necessary for comprehending the interplay between artificial intelligence (AI) and the prediction of stock prices.

Through the utilization of market microstructure theory, this study aims to provide a comprehensive understanding of how artificial intelligence (AI) can proficiently capture the influence of information distribution and trade complexities on price variations. Market Microstructure Theory explores how the trading processes and rules affect the price development and behaviour of financial markets. This theory offers a thorough comprehension of how the dissemination of information and intricacies in trade impact fluctuations in prices. Our study intends to showcase the effectiveness of AI in capturing these dynamics and improving stock price prediction using this theory.

Furthermore, the utilization of complex systems theory enhances our technique by emphasizing the dynamic and nonlinear characteristics of the stock market. In this context, the irreplaceable skills of artificial intelligence in pattern identification and predictive analytics are particularly relevant. The field of Complex Systems Theory focuses on highlighting the dynamic and non-linear nature of the stock market. This theory emphasizes the complex interaction of several market forces and their combined impact on market behavior. The hypothesis is relevant to our work since AI possesses an indispensable capacity to detect patterns and offer predictive analytics in very intricate settings. Through the incorporation of this theory, our objective is to authenticate our research methods and emphasize the efficacy of machine learning algorithms in properly forecasting stock values.

In addition to market microstructure and complex systems theories, the research is also based on Efficient Market Hypothesis and Random Walk theory.

The Efficient Market Hypothesis (EMH) states that stock prices include all accessible information, implying that market participants are unable to continually outperform the market since it promptly responds to news. The EMH framework classifies markets into three categories: weak, semi-strong, and strong efficiency. The Efficient Market Hypothesis (EMH) suggests that persistently outperforming the market is implausible. However, it does recognize the existence of temporary market inefficiencies, which can create opportunities for investors to take advantage of price discrepancies.

The Random Walk Theory posits that stock prices adhere to a stochastic trajectory and that future price fluctuations are fundamentally indeterminable. Every price alteration is unaffected by preceding alterations, highlighting the difficulty of accurately forecasting fluctuations in stock prices. This theory is consistent with the ideas of the Efficient Market Hypothesis (EMH), emphasising the significance of unpredictability in financial markets and adding to the wider discussion on market efficiency.

This theoretical investigation serves to confirm our research methodology and elucidate the efficacy of machine learning algorithms in accurately predicting stock values.

In order to reduce the potential negative impact of inaccurate forecasts, our study places a high importance on precision as the primary factor for assessing the effectiveness of our method. Precise and dependable predictions of stock prices are crucial in giving investors with vital information for making well-informed decisions. This research highlights the superior efficacy of machine learning algorithms compared to older methods in terms of accuracy and reliability. The objective of our study is to confirm this by showcasing the improved efficiency of machine learning models in navigating the complexities of financial markets. By doing thorough research and employing innovative algorithmic methods, our goal is to achieve substantial advancements in financial analytics, which will have a dramatic impact on investor strategies.

Utilizing machine learning techniques for stock price prediction has the potential to greatly transform financial forecasting. This investigation validates the higher precision and dependability of these algorithms in comparison to traditional approaches. Through the utilization of extensive historical datasets, machine learning has the ability to reveal patterns and connections that are not easily discernible through conventional analysis. This enables the identification of important prediction indicators such as momentum, liquidity, and volatility. These competencies enhance comprehension of market dynamics, facilitating more knowledgeable investment choices. This study promotes the growth of investment methods based on data analysis, enhancing financial analytics and providing vital information for investors aiming to maximise their profits.

The subsequent sections of this research study will provide a comprehensive examination of the existing literature, outline the research technique used, and analyze the study's conclusions. These findings will offer useful insights for investors and scholars, informing future endeavors in the field of financial forecasting.

LITERATURE REVIEW

The prediction of stock price movements is a complex task that has attracted considerable scholarly interest within the field of finance. There are three primary approaches that are commonly utilized in the forecast of stock prices: fundamental analysis, technical analysis, and statistical modeling. Predicting the future performance of the stock market is a significant challenge within the domain of the financial sector. The incorporation of modern technologies is crucial for producing viable results given the inherent unpredictability and hazards involved with the field. Qiu and Song (2016) assert that various factors exert influence on stock prices, including political matters, investor psychology, supply and demand dynamics, natural calamities, and other relevant determinants. The growing prominence of machine learning algorithms has spurred scholars to investigate their utilization for the purpose of improving the accuracy and effectiveness of predictions. Artificial intelligence has been recognized as a highly accurate methodology for predicting stock prices. The utilization of machine learning methodologies, such as Artificial Neural Network (ANN), Support Vector Machines (SVM), and Long Short-Term Memory (LSTM), facilitates the forecasting of stock market prices. Significant progress has been made in the field of artificial neural networks (ANN), particularly in the areas of financial markets and corporate applications. Artificial neural networks (ANN) are frequently utilized by investors as a dependable approach to effectively forecast stock market indexes. Chong et al. (2017) argue that deep learning possesses the ability to extract underlying theoretical attributes from data and reveal concealed nonlinear relationships, without necessitating the use of econometric assumptions or human experience.

Using machine learning algorithms to predict stock prices is the subject of a substantial corpus of literature. Other studies have investigated more recent models such as deep learning and reinforcement learning. It enables investors and businesses to evaluate stocks quantitatively in order to make more profitable decisions. Back propagation algorithm-based machine learning models outperform other neural network topologies for stock market index prediction (Sureshkumar & Elango, 2012).

In an investigation, Yu & Yan (2020) estimate stock prices using historical data and Deep Neural Networks (DNNs). The authors developed a deep learning framework for predicting stock prices using convolutional neural network (CNN) and long short-term memory (LSTM) artificial neural networks (ANN). The model consisted of four layers: input, CNN, LSTM, and output. CNN extracted local features from input data, whereas LSTM captured temporal dependencies. The proposed model outperformed conventional machine learning and deep learning algorithms on four stock datasets. Selvamuthu et al. (2019) It is possible to develop an algorithm for stock forecasting using a large amount of data and cutting-edge technology. This facilitates the purchase, sale, and selection of securities. Machine Learning and Data Mining share similarities, yet they differ in their respective approaches. Machine Learning focuses on the observation and analysis of novel algorithms and models, whereas Data Mining primarily focuses on the analysis of these methods and models (Sheth & Shah, 2023).

Mokhtari et al. (2021) evaluate the effectiveness of machine learning approaches in forecasting stock market trends in this research investigation. The authors undertake a critical evaluation of the merits and demerits pertaining to deep learning techniques, statistical models, and regression models. To guarantee accurate predictions, feature engineering, data pre-processing, and model selection are critical components to consider. Additionally, a critical analysis is performed regarding the disadvantages linked to the implementation of artificial intelligence in the domain of stock market forecasting. The aforementioned limitations pertain to difficulties associated with the comprehension of models and the possible concern of excessively fitting past data. The potential of Artificial Intelligence (AI) and machine learning to enhance the accuracy of stock market predictions is promising; however, additional research is required to address the inherent limitations of these technologies.

The utilization of machine learning algorithms for the purpose of stock market prediction is examined by Sheth and Shah (2023) in their scholarly investigation. The authors explore the significance of stock market forecasting and its potential to assist investors in making well-informed decisions. This paper investigates a number of research that have utilized machine learning methods, including decision trees, random forests, and support vector machines (SVMs), to make predictions about the stock market. The authors conduct a comparative analysis of the accuracy of several stock market prediction models and deduce that support vector machines (SVMs) exhibit the highest level of accuracy. Furthermore, the researchers analyze studies that predict the performance of the stock market by utilizing technical indicators and data derived from social media platforms. The authors subsequently provide an innovative algorithm that predicts stock prices by integrating technical indicators with sentiment monitoring of social media. Based on empirical evidence derived from real-world stock market data, the suggested approach demonstrates superior performance compared to standard machine learning algorithms. The authors' conclusion posits that the application of machine learning algorithms holds potential in anticipating stock market trends, hence offering investors reliable projections. According to the study conducted by Henrique et al. (2019), Machine learning possesses the capability to discern patterns, regularities, and disparities inside statistical data by grasping historical events, previous pricing trends, and other relevant factors. The artificial neural network (ANN), support vector machine (SVM), and long short-term memory (LSTM) are machine learning methods employed for the purpose of forecasting stock values. As previously said, the stock market exhibits a high degree of volatility and has the potential to generate substantial profits. The development of a reliable methodology that assists investors in making predictions and attains the highest level of accuracy is crucial for the purpose of minimizing annual investment losses.

In addition, Ravikumar and Saraf (2020) employ regression and classification algorithms to forecast the stock prices of companies. From 2015 to 2020, they gathered data on stock prices and economic indicators. Using regression techniques such as linear, Lasso, and Ridge regression and classification algorithms such as Random Forest, K-Nearest Neighbors (KNN), and SVM, the authors predicted the stock values of companies. Random Forest and SVM algorithms outperformed other regression and classification methods in predicting the stock values of companies. The authors reported that economic parameters including crude oil prices, interest rates, and inflation rates improved predictions. By combining financial and economic data, machine learning algorithms can predict stock prices, thereby assisting investors and speculators in making informed stock market decisions.

Moein Aldin et al. (2012) evaluated forecasts for the TEPIX stock price index. This study calculates the effectiveness of practical metrics such as Moving Average, RSI, CCI, MACD, etc. using stock price, volume, and interest rate data. The index of stock prices consists of the closing, high, and low prices. An ANN model with a three-layered input structure predicts stock prices. There are versions for input, concealed, and output. The output value version predicts indices of the stock market. Utilize Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root-Mean Square Error (RMSE) to validate results. The criteria combination (160, 30, 0.1, 0.039) maximized the commercial and analytic performance of the ANN model. It is precise 94% of the time.

Several studies have focused on the application of machine learning algorithms to stock price forecasting over the past several years. This category of algorithms includes neural networks, support vector machines, decision trees, random forests, and regression models. The ability of neural networks, which are included in this class of algorithms, to identify nonlinear relationships within the data has made them especially useful for predicting stock prices.

The utilization of convolutional neural networks (CNNs) and long short-term memory networks (LSTM networks) was observed in a study conducted by Yang et al. (2020). The researchers found that their model exhibited superior performance compared to traditional statistical models and alternative machine learning methods in terms of prediction accuracy. The authors of this article provide a system that utilizes deep learning to make predictions on the volatility of stock prices. The suggested approach involves the integration of two powerful deep learning models, namely convolutional neural networks (CNN) and long short-term memory (LSTM) networks. The Convolutional Neural Network (CNN) architecture is employed to extract salient features from the input data. These extracted features are subsequently inputted into the Long Short-Term Memory (LSTM) network for the purpose of forecasting forthcoming price variations. The researchers conducted empirical investigations using real-world stock market data and evaluated the suggested methodology against many established machine learning techniques. The findings of this study indicate that the proposed strategy exhibits superior performance compared to standard methods in terms of predictive accuracy. Furthermore, the authors conducted a sensitivity analysis to assess the impact of different parameters on the precision of their forecasts.

Extensive research has been conducted on the integration of artificial intelligence (AI) in the forecast of stock prices. Research has shown that AI techniques, specifically artificial neural networks (ANNs), are quite effective in forecasting stock values (Horák & Krulickú, 2019). Furthermore, the application of deep learning models, such as convolutional neural networks and reinforcement learning, has demonstrated encouraging results in predicting worldwide stock markets, particularly in situations where there is a scarcity of training data (Lee et al., 2019). According to Aldhyani and Alzahrani (2022), the utilization of deep learning algorithms has been important in the advancement of precise frameworks for the modeling of stock market prices. The progress of artificial intelligence in predicting stock market trends has led to the development of novel methodologies. According to Zarandi et al. (2012), the integration of artificial neural networks with fuzzy inference systems in hybrid models has demonstrated enhanced predictive capabilities. Moreover, advancements in artificial intelligence have resulted in the development of intelligent models based on ensemble learning, which effectively navigate the unpredictable and turbulent characteristics of financial markets (Wang, 2021). The utilization of AI-driven models has played a crucial role in improving the accuracy of stock market fluctuations, even in the presence of inherent uncertainties (ALI & SURI, 2022).

Shajalal et al. (2023) intend to utilize deep learning methodologies in an upcoming research endeavor to predict product backorders. A minority of the items contained in their dataset demonstrate instances of backorders, while the majority of the products remain devoid of such incidents. To mitigate the problem of skewed data, the researchers suggest employing a deep neural network configuration featuring numerous concealed layers. By employing convolutional and recurrent neural network layers, the model is capable of discerning and encapsulating temporal patterns, thereby extracting pertinent attributes. The authors' methodology is evaluated by calculating the F1-score, precision, recall, and accuracy. The experimental results demonstrate that the proposed methodology exhibits superior performance compared to existing state-of-the-art approaches in accurately forecasting product backorders, specifically when asymmetrical data is involved. The researchers have determined that their proposed methodology accurately predicts the occurrence of product backorders; thus, it demonstrates the potential applicability of the methodology to other datasets that possess an imbalance in the distribution of classes. In their recent publication, Liu et al. (2023) unveiled an innovative deep learning architecture called the Deep Residual Attention Network (DRAN), which was specifically designed to predict stock prices. The Dual-Residual Attention Network (DRAN) is an architectural design that combines residual learning and attention mechanisms within a convolutional neural network (CNN). By integrating financial news and technical indicators as inputs, the model is capable of producing stock price predictions encompassing both immediate and extended time periods. The authors trained and validated the model utilizing actual stock price datasets. Following this, the authors conducted a comparative analysis of the results obtained from their model in relation to three well-established machine learning models: the Long Short-Term Memory (LSTM) network, Support Vector Regression (SVR), and Random Forest Regression (RFR). The findings of the research demonstrate that the DRAN model demonstrated enhanced efficacy in forecasting stock prices with greater consistency and precision when compared to alternative models. The results of this study indicate that the DRAN model exhibits considerable potential for real-world implementations in the domain of stock price prediction.

Wanjawa (2016) utilized reference data acquired from the New York Stock Exchange (NYSE) and the Nairobi Securities Exchange (NSE) in order to generate forecasts regarding stock indices in their research. A feedforward complex perceptron was implemented, featuring error backpropagation, within a neural network architecture denoted as 5:21:21:1. During the training phase, which spanned 130,000 iterations, 80% of the available data was utilized. The Mean Absolute Percentage Error (MAPE) values that have been projected by the Nairobi Securities Exchange and New York Stock Exchange vary between 0.71 and 2.77 percent. Preliminary analysis unveiled a failure rate of 0.7% for the prototype. During the second evaluation, the new model was compared to two alternative models. Out of the models that were assessed, it was determined that the one with the least Root Mean Square Error (RMSE) was 1.83. For the preceding period, MAPE predicted a stock return of 0.71 percent. It was determined that the optimal model was 5:21:21:1. The implementation of the training procedure required the use of the most accurate training sets, comprising 1,000 records, or 80% of the total dataset. The training units underwent extensive iterations of training, totaling 130,000 repetitions.

Furthermore, a comprehensive analysis of feature importance was undertaken to ascertain the primary characteristics that exerted the most influence on the prediction of future stock values. Based on the available data, it is evident that specific technical indicators, namely the Relative Strength Index (RSI) and the Moving Average Convergence Divergence (MACD), hold greater significance compared to other indicators. Conversely, the impact of news item emotion on forecast performance is rather minimal. In general, the study introduces an innovative approach that utilizes deep learning techniques to forecast stock prices. This particular method exhibits promising promise in terms of its applicability to financial analysts and investors.

Although machine learning algorithms have demonstrated considerable efficacy, there are numerous challenges that must be addressed prior to their successful application in the domain of stock price prediction. The presence of financial data poses a substantial challenge due to its vulnerability to swift fluctuations and substantial levels of volatility. Furthermore, it is imperative to consider market trends, economic conditions, and geopolitical events as they possess the potential to impact stock values.

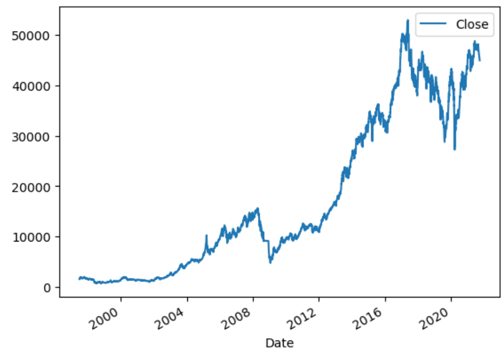
The collective outcomes of these studies indicate that machine learning algorithms have the potential to be advantageous in forecasting swings in stock prices. In order to enhance the precision of these models, it proves advantageous to amalgamate data from other sources, including social media platforms, technical indicators, and the sentiment expressed in news stories. This study employs a special methodology of optimizing the performance of a stock price prediction algorithm by maximizing accuracy as the error metric, which is considered a realistic technique.

METHODOLOGY OF STUDY

The primary objective of this research is to design a machine learning model that can accurately forecast the movement of stock prices. The focus of this study is to develop a model that not only maximizes profits but also minimizes potential risks associated with stock trading. The intended outcome is to make profitable trades by buying stocks when the market opens and selling them when the market closes, with a high level of confidence that the price will increase or decrease. We are using the KSE-100 index as our desired portfolio in which we are interested to invest.

The historical data of portfolio index price (KSE-100) is collected from Yahoo Finance ranging from 1997 to 2023. The downloaded dataset included the Open, High, Low, Close, Volume, Dividends and Stock Splits data. The close prices data is displayed in the line graph below. The KSE 100 is an index tracking the performance of the top 100 businesses traded on Pakistan's Karachi Stock Exchange (KSE). In this line graph, the time period is shown along the horizontal axis, which extends from 1997 to 2023. The value of the KSE 100 index is plotted against time on the vertical axis. If the line is going up or down, that means the market is doing well or poorly. If the line is rising, market performance is improving; if it is falling, market performance is deteriorating. The line chart of the KSE 100 may show some spikes and dips that signify temporary market movements. They may result from national or international economic situations, political unrest, or other factors. The line's long-term trend, however, reveals the market's cumulative performance over time.

Figure 1: Line Graphs of Close Price KSE-100



The dataset is organized to make projections about future stock prices based on previous stock prices. The dataset has been divided into two sections: the training data section and the test data section. The machine learning model is trained with the help of the training data, while the performance of the model is assessed with the assistance of the test data. Therefore, we need to prepare the data before we can make any predictions regarding whether or not there will be an increase in the price of the stock tomorrow based on the data from today. Our goal is to determine whether the price of the stock goes up or down the following day, which is represented by the numbers 1 and 0 correspondingly. We will move the data from the days before "forward" by one day so that we can utilize historical data to create forecasts. This will help us avoid making the typical mistake of using the same day's data for prediction, which is quite common. In the end, we finished off the process of creating our training data set by combining the shifted data with the target variable.

The methodology employed in this study aims to enhance the performance of the machine learning model through the integration of three major steps while employing the data analysis. Initially, the Sliding Window Methodology is employed, ensuring that the model is consistently refreshed with the most up-to-date and pertinent data. This methodology guarantees that our model maintains alignment with the changing dynamics of the market, consequently augmenting its capacity to predict trends with heightened accuracy and promptness. The Sliding Window Methodology is highly preferred because to its capacity to handle non-stationary time series data, which is a prevalent feature in financial markets where patterns and trends can change rapidly over time. This strategy effectively captures the most recent market behaviors and minimizes any delay that may arise from using static data sets by consistently updating the training data.

Subsequently, the process of backtesting is executed, which is a crucial stage in which the model is employed on past data. This assessment not only evaluates the model's ability to make accurate forecasts in various market circumstances but also identifies potential areas for improvement, thereby ensuring the reliability and robustness of our projections. Backtesting offers an in-depth review of model performance, enabling an analysis of its ability to make accurate predictions in various market situations. This approach is superior to basic in-sample testing because it allows for a better understanding of how the model would have performed in real-world situations, resulting in a more comprehensive assessment of its effectiveness.

Finally, we explore the process of hyper-parameter tuning, utilizing a comprehensive set of tools such as grid search, random search, Bayesian optimization, and evolutionary algorithms. The purpose of this rigorous procedure is to enhance the configuration of our Random Forest classifier, resulting in a significant improvement in the precision of our forecasts. Hyper-parameter tuning is crucial for maximising model performance, since it systematically explores the optimal parameters that control the learning process. Bayesian optimisation and evolutionary algorithms are especially advantageous methods because they are efficient in discovering optimal solutions in spaces with a large number of dimensions and they have the ability to balance exploration and exploitation, respectively.

The selection of these methods over alternative approaches is motivated by their proven effectiveness in addressing the special difficulties presented by financial time series data, including non-stationarity, high dimensionality, and the existence of noise. The Sliding Window Methodology's flexibility, comprehensive backtesting of real-world performance, and advanced hyper-parameter tuning techniques combine to create a machine learning model that demonstrates both precision and remarkable flexibility. This model is capable of providing insightful forecasts on stock price fluctuations in the Pakistan Stock Exchange.

These approaches are more advanced than traditional methods like static training datasets or manual parameter tuning since they provide a dynamic and iterative process that consistently enhances the model's performance. By ensuring the accuracy and robustness of our projections, we provide investors with a more dependable instrument that can withstand the constantly changing conditions of the stock market.

RESULTS AND INTERPRETATION

The process of training the model can then begin when the data has been partitioned into two sections. The act of teaching a machine learning model to recognize patterns in data and to make correct predictions or judgments based on those patterns is referred to as "model training." In order to set up the target, we start by making a new 'DataFrame' called 'data' and copying the 'Close' column into it. After that, we rename the column to 'actual_close' and save the changes. Because of this, we are able to monitor the real price at which the market closed each day. Next, we will build up the target by using the pandas rolling function to examine every two rows of the DataFrame. This will allow us to set up the target. This enables us to evaluate how the closing price of one day stacks up against the pricing of the following day. If the price at the end of the second trading day is higher than the price at the end of the first trading day, we will put a value of one in the Target column to indicate that the price increased. If the price at the end of the second day is lower than or equal to the price at the end of the first day, we put a value of 0 in the 'Target' column to indicate that the price went down. For the sake of this comparison, the 'Close' column will be utilized. The resulting 'Target' column has been updated to include the binary values that predicted by our machine learning model. The whole training process is discussed below;

Let X represent the historical data for the stock price, and let Y represent the target variable that indicates whether or not there is an increase or decrease in the stock price tomorrow (1 or 0, respectively). This can be expressed mathematically as follows:

$$X = \{x_1, x_2, x_3, \ldots, x_n\}$$

$$Y = \{y_1, y_2, y_3, \ldots, y_n\}$$

where n refers to the total number of data points in the collection.

In order to get the data ready for machine learning, we have to move the data from the days before "forward" by one day so that we are not making predictions based on the same-day data. This may be represented as:

$$X\_shifted = \{x_2, x_3, x_4, \ldots, x_n, x_{n+1}\}$$
$$Y\_shifted = \{y_2, y_3, y_4, \ldots, y_n, y_{n+1}\}$$

where X_shifted and Y_shifted represent the shifted data.

We may generate our own training dataset by combining the data that has been shifted with the variable that represents the target, and this dataset can be described as follows:

Training data:

$$Training\ data: \{(x_2, y_2), (x_3, y_3), (x_4, y_4), \ldots, (x_n, y_n)\}$$

To summarize, the following equation can be used to represent the process that was stated in the previous paragraph:

$Training\ data: \{(X\_shifted[i], Y\_shifted[i])\}$ for i = 1 to n-1

Training data: (X_shifted[i],Y_shifted[i]) for i = 1 to n-1, where X_shifted[i] represents the ith element in the shifted X data and Y_shifted[i] represents the ith element in the shifted Y data. n is the total number of elements in the data set.

After training the model, the first 10 values of the data are showed in the table, it clearly shown if the values are increasing, the target value is given as 1 and when market decreased, it reported as value of zero.

Table 1: Data Training

Date	**Actual_Close**	**Target**
1997-07-02	1618.15002	NaN
1997-07-03	1648.84998	1
1997-07-07	1691.68005	1
1997-07-08	1726.17004	1
1997-07-09	1778.51001	1
1997-07-10	1745.34998	0
1997-07-14	1794.64002	1
1997-07-15	1810.83997	1
1997-07-16	1812.78003	1
1997-07-17	1845.30005	1

Now in the next step, we are going to apply the shift method to the DataFrame at this time, which will move all of the rows "ahead" by one business day. Because of this, the prices that were listed for 1997-07-02 are now associated with 1997-07-03, and the positions of the prices listed after that have been moved up one row. The purpose of this is to ensure that we are accurately projecting future pricing based on historical data.

If we didn't adjust the data in this way, we would be forecasting prices for 07-03 using data from 07-03, which would be an inaccurate assumption. Instead, we need to make our price predictions for 07-03 using information from 07-02. If we do not adhere to this concept, our model might operate well during testing, but it might not work in the real world, where we do not have knowledge of tomorrow's pricing to guide our forecasts. If we do not adhere to this approach, our model might fail to work in the real world. After this adjustment in the model, here we are presenting the first 10 values of our resultant data.

Table 2: Shifting Data "Forward"

Date	Open	High	Low	Close
02/07/1997	NaN	NaN	NaN	NaN
03/07/1997	1593	1618	1590	1618
07/07/1997	1617	1649	1612	1649
08/07/1997	1674	1699	1674	1692
09/07/1997	1695	1726	1689	1726
10/07/1997	1727	1786	1711	1779
14/07/1997	1779	1784	1664	1745
15/07/1997	1757	1795	1747	1795
16/07/1997	1797	1830	1795	1811
17/07/1997	1808	1818	1803	1813

The above table clearly indicates that data has been shifted forward by one day, as mentioned earlier. The first row now contains NaN values, as there is no previous day's data to shift to this row. The other rows contain the prices and other information for each day.

The next thing we need to do is combine the target variable with the columns that will be used to predict the outcome. In order to accomplish this goal, we will use the join method on DataFrames. As soon as we joined all of our data together, we realized that in order to predict the objective variable, we are utilizing data from the day before. The columns "Close," "Volume," "Open," "High," and

"Low" are the ones that are retained to make a prediction about our objective. Table no 3 represents first 10 values of our trained data, after incorporating all the previous steps, and now data is ready for creating our machine learning model.

Table 3: Final Training Data for Machine Learning Model

Dates	Close	Target	Close	Open	High	Low
3 July	1649	1	1618	1593	1618	1590
7 July	1692	1	1649	1617	1649	1612
8 July	1726	1	1692	1674	1699	1674
9 July	1779	1	1726	1695	1726	1689

10 July	1745	0	1779	1727	1786	1711
14 July	1795	1	1745	1779	1784	1664
15 July	1811	1	1795	1757	1795	1747
16 July	1813	1	1811	1797	1830	1795
17 July	1845	1	1813	1808	1818	1803
21 July	1912	1	1845	1812	1845	1809

SETTING UP THE MACHINE LEARNING MODEL, RANDOM FOREST CLASSIFIER:

The following step is to apply machine learning techniques in order to develop a trustworthy model for predicting stock prices. We have also made our way here in order to use the random forest classifier for the same purpose. Because it can detect nonlinear correlations in the data and is reasonably resistant to overfitting provided that the appropriate parameters are used, a random forest classifier is an excellent choice as a default model for many applications. This is due to the fact that it may be used to classify the data.

The Random Forest Classifier is a method of machine learning that generates forecasts through the utilization of a forest of decision trees. The following econometric equation can be used to represent the Random Forest Classifier algorithm:

$$y = f(X, \theta) + \varepsilon$$

Where:

The binary target variable, denoted as "y," signifies the directional movement of market prices for the portfolio. A value of 1 indicates an increase in price, while a value of 0 indicates a decrease. The matrix "X" encompasses predictor variables, specifically historical returns of the portfolio. The vector "β" represents unknown parameters that the algorithm aims to estimate during the training process. Lastly, the error term "ε" signifies the discrepancy between the predicted values and the actual values. The variable y is a binary target variable that denotes the directional movement of market prices for the portfolio, indicating either a rise or decrease.

The random forest model undergoes training using the provided dataset, resulting in the generation of the function f(X, θ). This function is subsequently utilized to make predictions pertaining to the target variable. The ensemble consists of many decision trees, where each tree is built using a distinct random subset of predictor variables and a distinct random sample of the data. The cumulative sum of the outcomes from each tree is computed before to its utilization in the ultimate forecast.

The subsequent parameters employed in this equation are as follows: The parameter n_estimators is set to 100, representing the number of decision trees that constitute the ensemble. The parameter min_samples_split is set to 200, indicating the minimum number of samples required to split a node in a decision tree. Lastly, the parameter random_state is set to 1, serving as the random seed that guarantees reproducibility.

TRAINING THE MODEL

When we've finished configuring the model, we'll be able to train it using the most recent one hundred rows of the dataset. For the purpose of predicting the next price, we are use all of the data, with the exception of the most recent 100 rows. When working with data from a time series, it is imperative that you never utilize information from the future to try to understand the past.

The fit technique educates the model by training it to make accurate predictions using our predictors. The following is a representation of this process that may be made using econometric equations:

Let Y, the future price of the KSE-100 index, be the goal variable representing an increase or reduction in the future (1/0), and let X, the behavior of past prices, serve as the predictor variables. Let Y be the target variable. We have historical data for the past T time periods, which are designated as (Y_t, X_t) for t = 1, 2, ..., T. These data cover the entirety of the time period in question. We wish to train a random forest classifier model to predict the direction of Y for the last 100 periods, and we will use the 100 data that we have collected.

The model can be represented as:

$$Y_t = f(X_t) + e_t$$

where f(.) is the function that maps X to Y, and e_t is the error term at time t. The random forest algorithm estimates this function using an ensemble of decision trees.

To train the model, we use the fit method with the following input arguments:

$$model.fit(X[:T-100], Y[:T-100])$$

This trains the model on the previous observations of X and Y.

After training the model, we can use it to predict the direction of Y for the last 100 periods:

$$Y\_pred = model.predict(X[T-100:])$$

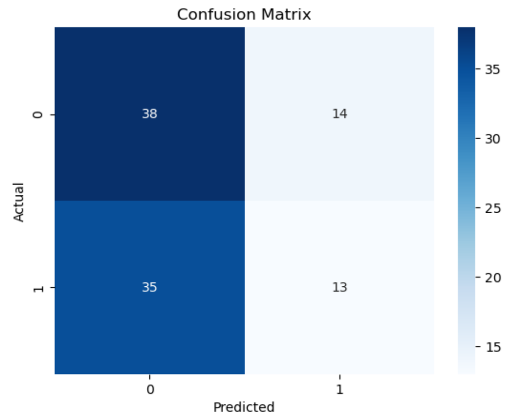
This predicts the direction of Y using the trained model on the last 100 observations of X.

CLASSIFICATION OF THE ACCURACY OF THE MODEL:

We utilized the model's precision as the error metric so that we could determine how accurate it was. The ratio of the overall number of true positives to the total number of positive predictions makes up the accuracy score. A high precision score indicates that the model accurately predicted positive values; in other words, when the model projected that the stock prices would go up, they actually went up. This is the same thing as saying that the model correctly forecasted negative values.

However, the current model has a precision score of 0.48, which indicates that the model's predictions are not very accurate and should not be relied upon. This indicates that the model has to be further refined because the stock prices only went up 48% of the time when the algorithm predicted that they would go up.

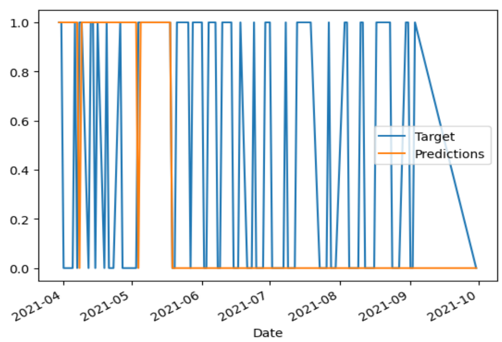
Figure 2: Confusion Matrix



The actual values are listed in the rows, while the expected values are listed in the columns. In light of the confusion matrix presented here, the model made the following predictions: 38 true negatives (TN), 13 true positives (TP), 14 false positives (FP), and 35 false negatives (FN). This indicates that 38% of the observations were correct, 14% of the observations were erroneous, 35% of the observations were false negatives, and 13% of the observations were correct.

It would appear that the model is not functioning up to expectations because it is only correct 48% of the time for the direction. We may look into this matter further by making a graph that compares the projected numbers to the actual values. The figure demonstrates that the model forecasted that the price would continue to increase each day. It does not make any predictions on the movement of the data. Therefore, despite the fact that this is not the best possible scenario, we need to make some adjustments to our model before moving on to the subsequent stage of backtesting.

Figure 3: Prediction Plot



On the other hand, this is merely the preliminary configuration of our model to check that everything is operating as it should. Backtesting our model using the full price history will allow us to derive an error metric that is representative of the true level of error. After we have finished putting the model together, the following step will be to do this.

Backtesting of the Model

It is vital to generate predictions on the whole dataset in order to improve our model rather than only focusing on the 100 most recent rows alone. By doing so, we are able to acquire an estimate of the model's inaccuracy that is more accurate; this is because the market conditions over the past 100 days may have been exceptional and may have an impact on future projections.

Backtesting, in which we only use data that was collected before the day we are attempting to predict, is something that needs to be done in order for this to be possible. It is not feasible to use data from the future to anticipate what happened in the real world in the past. Because of this, we must steer clear of using data collected after the day for which we are making a prediction.

Backtesting is accomplished by recursively going over all of the data in the dataset and training a model at intervals of 750 rows. Now that the process has been outlined, we will develop a function that will help us follow it, which will prevent us from having to rewrite the code each time we wish to run a backtest. We should train the model more frequently than every 750 rows, however for the sake of efficiency, we will increase this number. Idealistically, we should train the model more frequently than every 750 rows.

The result of performing backtesting and making adjustments to the hyper-parameters of the model yielded a new precession score of 58%. The previous score, which was just 48% accurate, has been significantly improved to 58%, representing a huge rise. Backtesting and hyper-parameter tweaking are two critical processes that must be completed in order to improve the overall performance of a machine learning model. Chong et al., (2017) & Yang et al., (2020), in their studies emphasize the use of advanced machine learning techniques, including deep learning models like CNNs and LSTMs, to enhance precision in stock price prediction. The researchers highlighted the iterative process of model refinement and parameter tuning to achieve better accuracy and effectiveness, which is consistent with our approach of backtesting and hyper-parameter adjustments leading to a significant increase in precision from 48% to 58%.

Following is the classification table after backtesting the model:

Table 4: Classification Table

Class	**Precision**	**Recall**	**F1-score**	**Support**
0	47%	76%	58%	2001
1	58%	28%	38%	2384
Total	53%	52%	48%	4385

This table shows the performance of the classifier on the test set. The precision for class 0.0 is 47%, which means that when the classifier predicts a stock will go down, it is correct

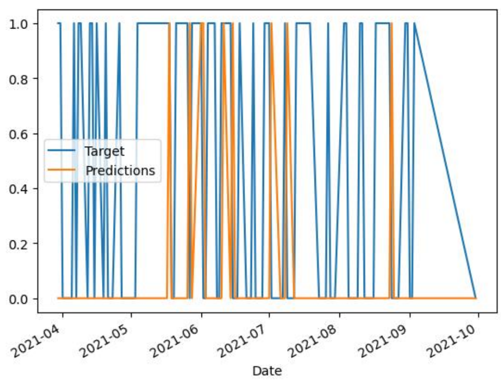
47% of the time. The recall for class 0.0 is 76%, which means that of all the stocks that actually went down, the classifier identified 76% of them correctly. The F1-score for class 0.0 is 58%, which is the harmonic mean of precision and recall for that class.

The precision for class 1.0 is 58%, which means that when the classifier predicts a stock will go up, it is correct 58% of the time. The recall for class 1.0 is 28%, which means that of all the stocks that actually went up, the classifier identified only 28% of them correctly. The F1-score for class 1.0 is 38%, which is the harmonic mean of precision and recall for that class. The study by Ravikumar & Saraf, (2020), which employs regression and classification algorithms including Random Forest and SVMs, shares similarities with our approach of evaluating precision and recall for different classes (stock increase and decrease). The researchers' demonstrate the importance of these metrics in assessing model performance accurately, which resonates with our findings of 47% precision for class 0.0 and 58% precision for class 1.0, albeit with varying recall rates.

Overall, the classifier has an accuracy of 50%, which is not very good. The macro average of precision, recall, and F1-score is around 52%, while the weighted average is around 48%. This suggests that the classifier is performing well on either class then the previous results, but although may need further improvements. Research by Mokhtari et al., (2021) discusses the iterative nature of optimizing machine learning models for stock market prediction. The researchers emphasize the need for continuous improvement through feature engineering, data preprocessing, and model selection, which parallels our efforts to achieve an overall classifier accuracy of 50% and optimize performance metrics like macro and weighted averages.

The precision score could be enhanced with the help of the number of factors, enhancement of the historical data, selection of the more stable market, inclusion of the more factors that determine the future price, selection of the more comprehensive technique and selection of more economic indicators.

Figure 4: Prediction Plot after Backtesting



Significant implications arise for investors and financial institutions that depend on precise and timely forecasts of stock valuations from the findings of this research. Investors can potentially enhance their investment decision-making process and potentially achieve greater returns by leveraging machine learning techniques to analyze historical data and forecast future prices.

Furthermore, this research offers significant contributions by shedding light on the possible implementations of machine learning in the domain of financial analysis and forecasting. With the advancement of artificial intelligence, the potential for applying machine learning algorithms to financial analysis continues to grow. It is expected that the utilization of this application will enhance the accuracy and efficiency of financial analysis.

However, it is crucial to bear in mind that machine learning models are flawed and cannot provide guaranteed accurate predictions in all circumstance. When deliberating on investment selections, it is imperative for investors to exhibit prudence and acknowledge that machine learning forecasts are merely one of several pertinent factors to be taken into account.

FUTURE DIRECTIONS

This study utilizes a Random Forest classifier, which is a commonly employed method in the field of machine learning. Nevertheless, the use of advanced techniques, such as neural networks, deep learning, and gradient boosting, presents an opportunity to significantly enhance the precision of the model. The aforementioned sophisticated algorithms demonstrate a high level of proficiency in identifying complex associations between the various features and the dependent variable.

The utilization of natural language processing techniques is of utmost importance in the field of sentiment analysis, specifically when analyzing news articles and social media content. Sentiment analysis encompasses the process of extracting and quantifying emotions, opinions, and attitudes from textual data, hence offering crucial insights into the market's assessments of particular assets. By proficiently utilizing sentiment analysis, individuals can greatly enhance the prediction capabilities of the model by acquiring a more comprehensive comprehension of the market's opinion towards specific assets.

The incorporation of alternate data sources is gaining prominence in scholarly discussions. These sources, which deviate from the customary data sets commonly used in traditional research methodologies, encompass a range of options such as satellite photography and web data scraping. The practice of web scraping has the potential to provide real-time data on consumer behavior and online transactions. Similarly, the utilization of satellite images can offer valuable insights on economic activity and the production processes of different commodities. The deliberate integration of these various data sources has the potential to enhance the precision of the model.

In relation to the existing feature set of the model, which is mostly based on technical indicators, it appears prudent to consider the incorporation of supplementary components. For example, the incorporation of price-volume correlations and volatility indicators, which may be derived from available data, may offer potential benefits. The incorporation of additional features through the process of feature engineering has the potential to offer a comprehensive portrayal of market dynamics, hence improving the accuracy of the model.

Finally, it is worth considering the concept of ensemble approaches. These methodologies entail the amalgamation of numerous prediction models in order to enhance the overall accuracy of forecasts. By utilizing the existing model in conjunction with several classifiers or regression models, it is possible to construct an ensemble that collectively enhances the accuracy of predictions. Ensemble approaches have demonstrated notable efficacy in mitigating the danger of overfitting and enhancing the model's generalization capacity.

CONCLUSION

The objective of this study is to investigate the utilization of machine learning techniques for the purpose of predicting stock prices. The study involves the development and training of a Random Forest classifier using historical data obtained from the Pakistan stock exchange. The aim of this study is to forecast future patterns in stock prices by assessing their potential for either upward or downward movement. The model utilizes various technical indicators, including moving averages and the Relative Strength Index (RSI), as its primary components. This paper explores the complexities of hyperparameter tuning, the process of doing backtesting, and the subsequent evaluation of model performance.

After completing the backtesting procedure, the results indicate that the Random Forest classifier has a precision score of 58%, which is an improvement over the previous score of 48%. However, the model's accuracy is only 58%, which still requires refinement. With the addition of more factors, the improvement of historical data, the selection of a more stable market, the inclusion of more factors that determine the future price, the selection of a more comprehensive technique, and the selection of more economic indicators, the precision score can be enhanced.

The practical implications of the findings from this study on the application of machine learning approaches for stock price prediction in Pakistan's financial sector are substantial. Our research showcases a new method for predicting stock fluctuations by using a Random Forest classifier to the KSE-100 Index. This provides investors and financial institutions with a more dependable tool to make well-informed decisions. The improved prediction accuracy of this model, as demonstrated by our thorough backtesting, indicates that investors can confidently optimize their portfolios, which could result in enhanced market efficiency and liquidity in Pakistan. Furthermore, our methodology, which involves adjusting hyper-parameters and utilizing the sliding window approach, offers a detailed plan for creating advanced financial research tools powered by artificial intelligence that are specifically designed to suit the distinct characteristics of Pakistan's stock exchange. These technological improvements are anticipated to promote a more data-centric investing environment, ultimately enhancing the expansion and stability of the nation's financial industry.

Our research contributes to the scholarly conversation surrounding the convergence of artificial intelligence and financial market forecasting in Pakistan. Our research is based on the theoretical frameworks of market microstructure theory and complex systems theory. This approach allows us to provide novel perspectives on the predictive capabilities of machine learning in contexts that exhibit information asymmetry, liquidity, and order flow. The significance of these theoretical frameworks in comprehending the behavior of the Pakistani stock market is emphasized by our findings, establishing a strong basis for future research in this field. This addition serves to enhance theoretical knowledge and foster a more nuanced examination of the determinants that impact fluctuations in stock prices. Consequently, it facilitates a more profound comprehension of market dynamics within the framework of emerging nations such as Pakistan.

Future directions for improving the model's accuracy are also provided, including the use of more advanced algorithms such as neural networks, deep learning, and gradient boosting; sentiment analysis of news articles and social media; the use of alternative data sources such as satellite imagery and web scraping; and feature engineering. At the conclusion of the article, the author reminds readers that machine learning algorithms are not foolproof and that investors must exercise prudence when making investment decisions.

Reference

Aziza, M., Suryadi, & Syuliswati, A. (2021). The Effect of Fundamental and Macroeconomic Factors on Agricultural Sector Company Stock Prices Listed in the Indonesia Stock Exchange on the 2016-2018 Period. *Proceedings of 2nd Annual Management, Business and Economic Conference (AMBEC 2020)*, *183*(Ambec 2020), 138–141. https://doi.org/10.2991/aebmr.k.210717.029

Chen, W. (2023). Application of Market Cycle Analysis and LSTM in Prediction of Stock Price Movements. *BCP Business & Management*, *38*, 856–861. https://doi.org/10.54691/bcpbm.v38i.3787

Chong, E., Han, C., & Park, F. (2017). Deep Learning Networks for Stock Market Analysis and Prediction. *Expert Systems with Applications*, *83*(April), 187–205. http://ac.els-cdn.com/S0957417417302750/1-s2.0-S0957417417302750-main.pdf?_tid=0d300a54-78da-11e7-ab02-00000aacb35f&acdnat=1501826538_c99481212aa82d83961ec6ff566751a4

Gursida, H. (2017). The Influence of Fundamental and Macroeconomic Analysis on Stock Price. *Jurnal Terapan Manajemen Dan Bisnis*, *3*(2), 222. https://doi.org/10.26737/jtmb.v3i2.324

Han, C., & Fu, X. (2023). Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model. *Frontiers in Business, Economics and Management*, *8*(2), 51–54. https://doi.org/10.54097/fbem.v8i2.6616

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, *124*, 226–251. https://doi.org/10.1016/j.eswa.2019.01.012

Liu, M., Sheng, H., Zhang, N., Chen, Y., & Huang, L. (2023). A New Deep Network Model for Stock Price Prediction. In Y. Xu, H. Yan, H. Teng, J. Cai, & J. Li (Eds.), *Machine Learning for Cyber Security* (pp. 413–426). Springer Nature Switzerland.

Mehtab, S., & Sen, J. (2020). *Stock Price Prediction Using Convolutional Neural Networks on a Multivariate Timeseries*. https://doi.org/10.36227/techrxiv.15088734.v1

Moein Aldin, M., Dehghan Dehnavi, H., & Entezari, S. (2012). Evaluating the Employment of Technical Indicators in Predicting Stock Price Index Variations Using Artificial Neural Networks (Case Study: Tehran Stock Exchange). *International Journal of Business and Management*, *7*(15), 25–34. https://doi.org/10.5539/ijbm.v7n15p25

Mokhtari, S., Yen, K. K., & Liu, J. (2021). Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning. *International Journal of Computer Applications*, *183*(7), 1–8. https://doi.org/10.5120/ijca2021921347

Nafia, A., Yousfi, A., & Echaoui, A. (2023). Equity-Market-Neutral Strategy Portfolio Construction Using LSTM-Based Stock Prediction and Selection: An Application to S&P500 Consumer Staples Stocks. *International Journal of Financial Studies*, *11*(2). https://doi.org/10.3390/ijfs11020057

Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., & Wills, G. (2020). A survey on machine learning for stock price prediction: Algorithms and techniques. *FEMIB 2020 - Proceedings of the 2nd International Conference on Finance, Economics, Management and IT Business*, *Femib*, 63–71. https://doi.org/10.5220/0009340700630071

Qiu, M., & Song, Y. (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. *PLOS ONE*, *11*(5), e0155133. https://doi.org/10.1371/journal.pone.0155133

Ravikumar, S., & Saraf, P. (2020). Prediction of stock prices using machine learning (regression, classification) Algorithms. *2020 International Conference for Emerging Technology, INCET 2020*, 1–5. https://doi.org/10.1109/INCET49848.2020.9154061

Ruan, J., Wu, W., & Luo, J. (2020). Stock Price Prediction under Anomalous Circumstances. *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, 4787–4794. https://doi.org/10.1109/BigData50022.2020.9378030

Safari, A., & Badamchizadeh, M. A. (2023). Stock index forecasting using DACLAMNN: A new intelligent highly accurate hybrid ACLSTM/Markov neural network predictor. *Cognitive Computation and Systems*, *July*, 181–194. https://doi.org/10.1049/ccs2.12086

Selvamuthu, D., Kumar, V., & Mishra, A. (2019). Indian stock market prediction using artificial neural networks on tick data. *Financial Innovation*, *5*(1). https://doi.org/10.1186/s40854-019-0131-7

Shajalal, M., Hajek, P., & Abedin, M. Z. (2023). Product backorder prediction using deep neural network on imbalanced data. *International Journal of Production Research*, *61*(1), 302–319. https://doi.org/10.1080/00207543.2021.1901153

Sheth, D., & Shah, M. (2023). Predicting stock market using machine learning: best and accurate way to know future stock prices. *International Journal of System Assurance Engineering and Management*, *14*(1), 1–18. https://doi.org/10.1007/s13198-022-01811-1

Sureshkumar, K. K., & Elango, N. M. (2012). Performance Analysis of Stock Price Prediction. *Global Journal of Computer Science and Technology*, *12*(1), 19–26. https://computerresearch.org/index.php/computer/article/view/426/426

Wanjawa, B. W. (2016). *Evaluating the Performance of ANN Prediction System at Shanghai Stock Market in the Period 21-Sep-2016 to 11-Oct-2016*. *147*(March), 11–40. http://arxiv.org/abs/1612.02666

Yang, C., Zhai, J., Tao, G., & Haajek, P. (2020). Deep Learning for Price Movement Prediction Using Convolutional Neural Network and Long Short-Term Memory. *Mathematical Problems in Engineering*, *2020*. https://doi.org/10.1155/2020/2746845

Yu, P., & Yan, X. (2020). Stock price prediction based on deep neural networks. *Neural Computing and Applications*, *32*(6), 1609–1628. https://doi.org/10.1007/s00521-019-04212-x

Annex-A: Predicting Stock Prices through Machine Learning (Summary of Literature Review)

Authors	**Sample**	**Period**	**Frequency**	**Research Methodology/ Tool**
Sureshkumar, K.K. Elango, N.M. (2012)	TCS (India Stock Exchange)	2009-2011	Daily	Artificial Neural Network (ANN)
Moein Aldin, Mahmood Dehghan Dehnavi, Hasan Entezari, Somayye (2012)	Tehran Exchange Price Index (TEPIX)	2002-2009	Daily	Artificial Neural Network (ANN)
De Tre, Guy Hallez, Axel Bronselaer, Antoon (2014)	04 stocks (IBM, Dell Corporation, British airways and Ryanair)	2002-2005	Daily	Fuzzy multiagent system (FMAS)
Wanjawa, Wamkaya B (2016)	Shanghai Stock Market	Sept-Oct 2016	Daily	Artificial Neural Network (ANN)
Qiu, Mingyue Song, Yu (2016)	Nikkei 225 index	2007-2013	Daily	Artificial Neural Network (ANN)
Chong, Eunsuk Han, Chulwoo Park, Frank C.(2017)	stock returns of 38 stocks from the KOSPI market (South Korea)	2010-2014	Every five minute	Deep Neural Network (DNN)

Lee, Jinho Kim, Raehyun Koh, Yookyung Kang, Jaewoo (2019)	Data of 31 countries	2006-2017	Daily	Deep Q-Network JINHO
Horák, Jakub Krulický, Tomáš (2019)	UniPetrol (Czech Republic)	2006-2018	Daily	Artificial Neural Network (ANN)
Selvamuthu, Dharmaraja Kumar, Vineet Mishra, Abhishek (2019)	Reliance Private Limited (India)	30 NOV 2017 to 11 JAN 2018 (430,000 data points)	Tick data	Artificial Neural Network (ANN)
Ravikumar, Srinath Saraf, Prasad (2020)	Companies of S&P 500	2010-2017	Daily	Support Vector Machine, Logistic Regression, Decision Tree and Random Forest Classification
Yu, Pengfei Yan, Xuesong (2020)	Six stock indices (S&P 500, the Dow Jones industrial average (DJIA), the Nikkei 225 (N 225), the Hang Seng index (HSI), the China Securities index 300 (CSI 300) and the ChiNext index)	2008-2017	Daily	Deep Neural Network (DNN)
Yang, Can Zhai, Junjie Tao, Guihua Haajek, Petr (2020)	11 stock indices (CAC40, DJIA, S&P 500, NAS-DAQ, DAX, FTSE, NYSE, HSI, N225, SSE, and RUSSELL).	2010-2017	Daily	Convolutional neural network (CNN) long short-term memory (LSTM)
(Hindrayani et al., 2020)	4 stocks of Indonesia Stock Exchange	2016-2020	Daily	Decision Tree Regression
Mokhtari, Sohrab Yen, Kang K. Liu, Jin (2021)	AAPL (from Yahoo Finance)	2010-2021 & Sentiment based data of 6000 Tweets	Daily	ANN, DT, and SVM models, logistic regression (LR), Gaussian naive Bayes (GNB), Bernoulli Naive Bayes (BNB), random forest (RF), k-nearest neighbor (KNN), and XGboost (XGB) , linear regression and long short-term memory (LSTM)

Wang, Zijun (2021)	Shanghai composite index	2005-2019	Daily	Logistic Regression, Support Vector Machine
(J. Yang et al., 2022)	3 stocks of NYSE/NASDAQ	2012-2020	Daily	LASSO-LASSO, LASSO-LSTM
Aldhyani, Theyazn H.H. Alzahrani, Ali (2022)	Tesla, Inc., Apple, Inc. Data	2014-2017	Daily	long short-term memory (LSTM) and a hybrid of a convolutional neural network (CNN-LSTM)
Gülmez, Burak (2023)	DJIA	2018-2023	Daily	Long short-term memory (LSTM)
(Dahal et al., 2023)	S&P 500	2008-2021	Daily	Long short-term memory (LSTM) and gated recurrent unit (GRU)
(Li et al., 2023)	322 stocks of 11 S & P Sectors	2005-2017	Daily	Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU)
(Qi, 2023)	Stock of Shanghai stock market	2021-2022	Daily	Support Vector Machine
(Islam et al., 2023)	Saudi Dairy and Aramco Company	2020	Daily	Linear regression, decision tree, and support vector machine.
(Tekin et al., 2023)	Tesla	2010-2020	Daily	Principal Component Analysis (PCA)
(Sohrabi et al., 2023)	Tehran Stock Exchange	2020-2022	Daily	Long Short-Term Memory (LSTM)
(Zhang, 2023)	Apple, Google & Amazon stocks	2013-2022	Daily	Convolutional neural network (CNN)