Utilization of Machine Learning for predicting the Stock Price of Amazon

Dr.S.Selvakani
Assistant Professor and Head, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamilnadu, India – 631051.
Mrs.K.Vasumathi
Assistant Professor, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamilnadu, India – 631051.
Mrs.S.Tamilselvi
PG Scholar, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamilnadu,India – 631051.

----------ABSTRACT----------

The previous dataset can be utilized to obtain the current prediction of price. The function of predicting stock prices involves determining the future price of a company's stock or other traded commercial instrument. The stock market and exchanges offer significant financial gains, which attract investors. Although the prediction of stock prices is a challenging issue, it can be resolved by implementing various techniques. We propose to gather input data for this project, remove unwanted null values, extract relevant data for feature engineering, and then apply Decision tree and Linear regression models for training purposes. The stock market is characterized as dynamic, unpredictable, and non-linear in nature, making it difficult to predict stock prices. Several factors, including but not limited to political conditions, global economy, company financial reports, and performance, contribute to the challenge of predicting stock prices.

Keywords: Prediction, Machine Learning, Data Set, Stock Price prediction, Decision tree algorithm, and linear regression.

I. INTRODUCTION:

Stock price prediction involves estimating the current market value of a product. There are two primary functions of stock price prediction: historical analysis of the past 12 months and forecasting for the next 12 months. Sequential data processing techniques, such as Decision tree and Linear regression, are commonly used for this purpose. Stock prices can be predicted in three different ways: Regular price, Least possible price, and Best possible price.

Decision tree and Linear regression are two well-known techniques used for processing sequential data. The prediction of financial time series is an intricate task due to the noisy, non-stationary, and chaotic nature of the data. This is because financial time series data is highly unpredictable, and small variations in data can lead to significant changes in the outcome.

With the rapid development of information technologies in the last two decades, new techniques have emerged, inspired by the inter-connected neurons in the human brain. These techniques, commonly referred to as Artificial Neural Networks (ANNs), have been proposed as a solution to improve the prediction performance of time series data. ANNs are a powerful tool for modeling complex non-linear relationships between variables and are commonly used for financial time series prediction.

The use of ANNs has shown significant promise in improving prediction accuracy, and they have been used to predict stock prices, foreign exchange rates, and commodity prices, among others. By using ANNs, we can identify patterns in historical financial data and use them to make informed predictions about future prices. This can be extremely beneficial for investors who rely on accurate financial predictions to make informed decisions.

The prediction performance of the Autoregressive Integrated Moving Average (ARIMA) model is often compared against that of other advanced prediction models such as the Long Short-Term Memory (LSTM) and Gated-Recurrent Unit (GRU) for each BIST price index. Stock investment is widely recognized as one of the main investment tools in the financial markets.

In the field of finance, traditional and machine learning methods are frequently utilized to improve prediction precision, thereby assisting investors in avoiding risks and achieving better performance. These methods use

historical financial data to identify patterns and trends in the market, which are then used to make informed predictions about future market movements.

Machine learning techniques such as ANNs, Support Vector Machines (SVMs), and Random Forests (RFs) have shown great promise in improving prediction accuracy in finance. The use of these techniques can help investors make better-informed decisions and minimize their investment risks.

However, the accuracy of these methods depends on the selection of appropriate features from the raw data. Future studies are therefore planned to focus on extracting various uncorrelated features from the financial data and studying their multivariate structure. This will allow us to identify additional factors that influence market trends, which may not be immediately apparent from a univariate analysis.

The stock market is characterized as a dynamic, unpredictable, and non-linear system. Predicting stock prices is a challenging task because it depends on various factors such as political conditions, global economy, company's financial reports, and performance, among others. Machine learning techniques have been found to improve efficiencies in this area by 60-86 percent compared to traditional methods. While most previous work in this field has utilized classical algorithms like linear regression and linear models like Autoregressive Moving Average (ARMA) for predicting stock prices, recent studies have demonstrated that machine learning can significantly enhance stock market prediction.

Various machine learning techniques have shown promising results in predicting stock prices, including Support Vector Machine (SVM), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM). The stock market is known to be dynamic, unpredictable, and non-linear in nature, which makes predicting stock prices a challenging task. Traditionally, two main approaches have been proposed for predicting the stock price of an organization. However, machine learning techniques have the potential to significantly improve prediction accuracy and reduce the variance in the model. Therefore, using machine learning techniques for stock closing price prediction is a promising research direction.

It has the ability to perform both regression and classification tasks by combining multiple decision trees to determine the final output instead of relying on individual trees. In the stock market, every buyer and s eller tries to predict price movements to maximize profits and minimize losses. Artificial Intelligence (AI) has been applied to financial investment and has attracted extensive research. This paper presents a systematic review of the literature on AI applied to investments in the stock market, based on the dynamic, unpredictable, and nonlinear nature of the stock market. Machine learning algorithms such as Random Forest and Support Vector Machines have been used to predict stock values. We proposed a system called "Stock Market Price Prediction" which uses the Random Forest algorithm to predict stock prices. Predicting the stock market requires the ability to predict the effects of recent events on investors and construct complex nonlinear relationships bet ween input and output data.

The application of machine learning algorithms like Random Forest and Support Vector Machines has been widely explored for predicting stock values. In our proposed system, named "Stock market price prediction," we have implemented the random forest algorithm to accurately forecast the stock market price. It is important to note that stock market prediction requires the ability to anticipate the impact of recent events on the investors. By analyzing the stock market and establishing complex nonlinear relationships between input and output data, researchers aim to improve the precision of stock market predictions. This project aims to enhance the accuracy of stock market predictions by utilizing stock values as input data, ultimately benefiting investors by providing valuable insight into future market conditions.

II. REVIEW OF LITERATURE:

According to a study by O.E. Yigit, A.L.P Selcuk, and O.Z. Ersoy [4], financial time series prediction is a difficult task due to the series' noisy, non-stationary, and chaotic nature. However, with the rapid advancements in information technologies over the past two decades, neural network models inspired by the inter-connected neurons of the human brain have been proposed to enhance time series prediction performance. Specifically, the study compared the prediction performances of ARIMA, Long Short-Term Memory, and Gated-Recurrent Unit for each BIST price index.

According to O.E. yigit, A.L.P selcuk, O.Z.Ersoy [4], they presented some advantages, including:

- The paper discusses the concept of using Long Short-Term Memory for stock prediction, both for the short and long term.
- Real-world data is used to demonstrate the time series behavior and forecasting, which is crucial for making scientific, commercial, industrial, or economic decisions based on past observations.
- LSTM is an upgraded version of Recurrent Neural Network (RNN).

- Important studies in time series prediction have been conducted in the financial literature that focuses on deep learning.
- The diagnostic checking step is performed to determine if the derived model is adequate for forecasting by handling the residuals generated by prediction.

According to A. Kumar [5], the stock market is characterized by its dynamic, unpredictable, and non-linear nature. Predicting stock prices is a challenging task, as it depends on various factors, including but not limited to political conditions, global economy, and a company's financial reports and performance. Machine learning techniques have shown to improve efficiencies in this area by 60-86 percent compared to past methods. Most previous work in this field used classical algorithms like linear regression and some linear models like Autoregressive Moving Average (ARMA) for predicting stock prices.

A. Kumar [5] highlights the advantages of describing the idea of every Closing Price Prediction using Machine Learning Techniques in this paper. Thus, to maximize profit and minimize losses, predicting stock values in advance by analyzing trends over the last few years could prove highly useful in making stock market movements.

According to N. Agarwal and S. Gupta [3], machine learning algorithms such as Random Forest and Support Vector Machines have been utilized to predict stock values. The authors proposed a system called "Stock market price prediction" and utilized the Random Forest algorithm to predict the stock market price. Predicting the effect of recent events on investors is a crucial aspect of stock market prediction.

N. Agarwal and S. Gupta [3] stated the following advantages: • Every buyer and seller aims to predict stock market price movements to maximize profits and minimize losses. • Utilizing advanced technologies such as AI can enhance stock price prediction accuracy. • The problem was mainly tackled using two machine-learning libraries, namely scikit and numpy, for real analysis and prediction. • Moreover, investors can benefit significantly from the analysis by gaining insight into future market conditions.

Mahomet UgurGudeleka and Omer BeratSezera [1] aimed to give an overview of the latest DL models used in finance applications, as an increasing number of such models have emerged in conferences and journals. The paper presents different implementations of these models, such as Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM), to provide a starting point for researchers and practitioners in the field. • Deep Learning is a type of Machine Learning that utilizes multiple layers of Artificial Neural Networks (ANNs) to provide high-level data modeling. • The paper also highlights the ongoing research on DL model development, and the next section will provide more details on the implementation of these models in different areas, along with the preferred DL models.

Z. Hu, Y. Zhao and M. Khushi [6] have stated that there are various factors that contribute to the fluctuations in stock prices, such as macro-economic conditions, market trends, and investors' confidence in the company's management and performance. Forex is a globally significant financial market, and predicting exchange rates can aid investors in making informed decisions to increase returns and decrease risks..

• Z. Hu, Y. Zhao and M. Khushi [6] state that stock and foreign exchange (Forex) price prediction has always been a lucrative and popular area of research. • This paper provides a review of stock prediction, covering the dataset, variables, and models used. • The complex and nonlinear relationships in stock and Forex prediction can be fully captured by deep learning models. • Given the excellent performance of deep learning models in other research areas, it is feasible to predict stock and Forex trends using deep learning.

W. Jiang [7] aims to provide an updated review of recent research on deep learning models for stock market prediction. The survey will focus on the latest emerging deep learning, which is characterized by various structures of deep neural networks. Deep learning models have a strong ability to handle big data and learn the nonlinear relationship between input features and prediction target, resulting in better performance compared to both linear and machine learning models for stock market prediction tasks.

According to W. Jiang [7], stock market prediction is a classic problem that lies at the intersection of finance and computer science, and there are advantages to using technical indicators as handcrafted input features for machine learning models. In particular, deep learning models have shown better performance than both linear and traditional machine learning models on tasks related to stock market prediction. This survey will focus on the latest emerging deep learning structures, specifically recurrent neural networks, such as LSTM networks, for predicting stock market prices movements. The model's financial performance will also be validated by comparing it to some simple but valid investments. Despite the inherent complexity and dynamism of stock markets, there has been an ongoing debate about the predictability of stock returns.

D.M.Q. Nelson and A.C.M. Pereira [2] discuss the concepts of stock price movement prediction in this article, both in the context of stock markets and machine learning. They note that Long Short Term Memory (LSTM) networks, which are a type of deep and recurrent neural network, are used in this project for predicting stock market price movements. Despite the inherent complexity and dynamism of stock markets, there has been an ongoing debate about the predictability of stock returns. While LSTM networks have shown notable performance in Natural Language Processing (NLP), they have been mainly used with news text data as input to predict price trends and stock indexes, to determine whether a stock price will increase, decrease or remain the same on a given day.

According to D.M.Q. Nelson and A.C.M. Pereira [2], using Machine Learning algorithms to learn from historic price data is a common approach for predicting future stock prices. They note that compared to other recurrent neural networks that can only memorize short sequences, these algorithms are more capable of handling long input sequences. The model's financial performance is also validated by comparing it to some simple but valid investments. Despite the inherent complexity and dynamism of stock markets, there has been a constant debate about the predictability of stock returns.

III. PROPOSED METHODOLOGY:

The proposed system involves the availability of opinions in numerous sources, including blogs, social networking sites like Facebook and email, news portals, and e-commerce sites. The objective of the paper is to predict future stock prices of Amazon using various data mining techniques. The authors plan to utilize several data mining algorithms and suggest different methods for combining the results obtained from them. The algorithms employed in this study include decision tree and linear regression.

Linear regression:

Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), consequently called linear regression*as shown in Figure 1.*

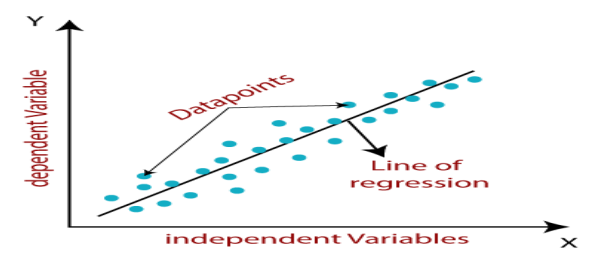

Figure 1: Linear Regression

IV. EXPERIMENTALRESULTS:

In this study, we proposed two different approaches to address the data and evaluated their performance on a CNN-based model for predicting the next day's trade action (buy, sell, or hold) of Dow30 stocks. The dataset used in the study was obtained from finance.yahoo.com and comprised of daily price data for Dow30 stocks from 01/01/2008 to 15/12/2021. The adjusted closing price data was used to label the images as buy, sell, or hold, while the other price data such as close, high, low, and volume were utilized for labeling the datasets in both approaches. We utilized a sliding window algorithm for labeling the datasets in this study.

Algorithm 1:

The dataset was labeled using the proposed labeling algorithm. Each trading day in the dataset was assigned one of three labels: "Buy", "Sell", or "Hold". The following rules were applied to the data from the past five days:

- If there is only one "Buy" and no "Sell" in the label values of the past five days, the data is labeled as "Buy".
- If there is only one "Sell" and no "Buy" in the label values of the past five days, the data is labeled as "Sell".
- If the label values of the past five days are either ['Buy', 'Hold', 'Hold', 'Sell'] or ['Buy', 'Hold', 'Hold', 'Sell'], the data is labeled as "Sell".

Feature selection:In order to predict the stock prices for the next day, it is crucial to identify the input variables that the model will use. To accomplish this, it is necessary to analyze the factors that impact stock prices.

Image creation: Through our experiments, we discovered that the most effective approach is to consider the technical indicators and daily closing prices of gold and oil from Yahoo Finance over the last five days. To construct a matrix that represents daily prices and their associated features, we selected a period and calculated the values of 15 technical indicators for 7, 14, and 21-day periods. Using these values, we created a 3 x 15

matrix for one trading day. We repeated this process for the last five days and merged the resulting matrices to create a 15 x 15 matrix. Finally, we transformed this matrix into an image.

CNN Model: To evaluate the efficacy of labeling and feature selection methods, we utilized the CNN architecture proposed by SezerandOzbayoglu in this study. This CNN architecture comprises of 8 layers, including an input layer, two convolutional layers, a max drop outs layer, a fully connected layer, and an output layer. The first two convolution layers employ a filter size of 3 x 3, with 32 and 64 filters respectively.

Performance measure: The confusion matrix frequently used to evaluate the model's performance within thescope of classification problems. The information included in this matrix refers to follows:

- **True Positive (TP):** It is the number of samples thatare actual class value is ''positive'' and labeled the''positive'' class as a result of the classification.
- **False Positive (FP**): It is the number of samples thatare actual class value is ''negative'' and labeled the''positive'' class as a result of the classification.
- **True Negative (TN)**: It is the number of samples thatare actual class value is ''negative'' and labeled the''negative'' class as a result of the classification.
- **False Negative (FN):** It is the number of samples that areactual class value is ''positive'' and labeled the ''negative'' class as a result of the classification.

Accuracy Formula:

accuracy = (TN+TP)\ (TN+TP+FN+FP)

recall = TP \ (TP + FN)

Proposed model:

		Proposed Model
Accuracy(%)		**80.23**
AAPL	**Accuracy(%)**	**77.03**
	F1 score	**0.76**
GS	**Accuracy(%)**	**77.03**
	F1 score	**0.77**

precision= TP \ (TP + FP**)**

F1 Score= 2 .precision .recall \ (precision + recall)

V. CONCLUSION:

Our research has demonstrated that our Amazon stock price prediction model exhibits higher accuracy and a better fitting degree compared to the self-coding network. Moving forward, we aim to enhance our model by introducing more relevant factors, with the goal of improving the arbitrage pricing model's ability to predict Amazon's unique circumstances. Through further analysis of Amazon's industry characteristics and market position, we have gained a better understanding of the complex relationship between various factors and changes in stock prices, and have realized that this relationship varies across different companies. Although the self-coding algorithm did not produce better results, the coefficients derived from it serve as evidence supporting the reliability of the impact coefficients of various systemic factors on stock prices, which are deeply described by the arbitrage pricing model.

REFERENCES.

1. A. M. Ozbayoglu, M. U. Gudelek, and O. B. Sezer, ''Deep learning for financial applications: A survey,'' Appl. Soft Comput., vol. 93, Aug. 2020.

2. D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, ''Stock market's price movement prediction with LSTM neural networks,'' in Proc. Int. Joint Conf. Neural Netw. (IJCNN), May 2017.

3. I. Parmar, N. Agarwal, S. Saxena, R. Arora, S. Gupta, H. Dhiman, and L. Chouhan, ''Stock market prediction using machine learning,'' in Proc. 1st Int. Conf. Secure Cyber Comput. Commun. (ICSCCC), Dec. 2018.
4. L. P. Selçuk, Ö. E. Yiğit, and Ö. Z. Ersoy, ''Prediction of bist price indices: A comparative study between traditional and deep learning meth-ods,'' Sigma J. Eng. Natural Sci., mar 2020.

5. M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, ‘‘Stock closing price prediction using machine learning techniques,’’ Proc. Comput. Sci., Jan. 2020.
6. T. Gao, Y. Chai, and Y. Liu, ‘‘Applying long short term momory neural networks for predicting stock closing price,’’ in Proc. 8th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS), Nov. 2017.
7. W. Jiang, ‘‘Applications of deep learning in stock market prediction: Recent progress,’’ Expert Syst. Appl., vol. 184, Dec. 2021.