Stock Price Prediction Using Machine Learning*

Aniket Gupta
Department of CSE
Galgotias University
Greater Noida, Uttar Pradesh, India
aniketgupta41099@gmail.com

Rishabh Chaudhary
Department of CSE Galgotias University
Greater Noida, Uttar Pradesh, India
chaudharyrishav30@gmail.com

Dr.Medhavi Malik
Assistant Professor,Department of CSE Galgotias University
Greater Noida, Uttar Pradesh, India

***Abstract*—The stock market has always offered investors a chance to make money, but the majority of profits are generated by analysis of the present and historical market conditions, followed by proactive measures. There are several factors that need to be taken into account before making decisions due to the current overheated market economy.**

a stock market deal that is profitable. Analyzing all of these variables and influencing factors manually would be laborious and prone to mistakes. Therefore, the examination of such an apparently chaotic system is best served by a machine learning approach. In this project, machine learning is being used to predict many features of a specific stock or index, including future values of the opening price, closing price, index value, etc.

***Keywords*—Keywords—Stock Market Predict Price; Machine Learning**

I. Introduction

The act of attempting to predict the future value of a company shares or other securities on the stock market a financial product traded on a market. The stock market is a significant component of the national economy and is essential to the development of the nation's industry and commerce, both of which have an impact on the national economy.Both investors and business are interested in the stock market and want to know whether a particular stock will increase or decrease over a specific time period. The stock market serves as a company's main source of capital for growing its operations.It is predicated on the idea of supply and demand. When there is greater demand for a company's stock, the price of its shares rises; conversely, when there is less demand for a company's stock, the price of its shares falls. Due to the engagement of numerous industries and businesses, it contains extremely vast sets of data from which it is challenging to manually extract information and analyse their work trends. The study and forecasting of the stock market will show market trends and indicate when to buy stocks. A stock's future price prediction that is accurate could result in a sizable profit. This is accomplished by representing different conditions with big historical market data and confirming that time series patterns have statistically significant forecasting potential for high probabilities of profitable trades and high profitable returns for the competitive company investment. A data analysis technique called machine learning automates the creation of analytical models. Machine learning enables computers to discover hidden insights without being explicitly instructed where to seek by using algorithms that iteratively learn from data. Although the application of machine learning algorithms is typically focused on technical analysis, it can be advantageous to incorporate fundamental analytical techniques.This study describes the numerous approaches used to apply machine learning to stock forecasting and makes new, powerful suggestions for future development. Many intriguing studies have been conducted recently in the field of using machine learning algorithms to analyse price trends and forecast changes in stock prices and indexes. The majority of stock traders today rely on intelligent trading systems to predict prices based on various circumstances and variables, enabling them to make quick investment decisions.

A. Literature Survey

The last few decades have seen a rise in interest in the field of research surrounding stock return forecasting. The majority of the time, the researchers tried to find a linear connection between the stock returns and the input macroeconomic data. Many books on nonlinear statistical modelling of stock returns have been written since nonlinearity in stock market index returns was discovered; nevertheless, most of them called for the nonlinear model to be stated before the estimation was carried out.The study of stock return forecasting has attracted more attention during the past few decades. The researchers often looked for a linear relationship between the stock returns and the input macroeconomic data. Since the discovery of nonlinearity in stock market index returns, many publications on nonlinear statistical modelling of stock returns have been written; nevertheless, the majority of them required the nonlinear model to be described before the estimation was done.While some researchers investigated the incorporation of diverse market data and macro-economic variables, others utilised input data from a single time series. Before supplying these input data sets to the ANN for predicting, some researchers even pre-processed them.

B. Future Scope

Due to constantly fluctuating stock values that depend on numerous factors and generate complicated patterns, predicting stock market returns is a difficult process. Only a few parameters, such as high, low, open, close, adjacent close value of stock prices, volume of shares traded, etc., are included in the historical dataset that is accessible on the firm website, and these features are insufficient.Using the already-existing factors, new variables have been developed to achieve a higher degree of accuracy in the forecasted price value. The stock's closing price the following day is forecasted using ANN, and RF is also utilised to do a comparative study. In comparison to RF, ANN provides better stock price prediction, according to a comparative analysis based on RMSE, MAPE, and MBE values. Results indicate that RMSE (0.42), MAPE (0.77), and MBE are the best values produced by the ANN model (0.013).For further work, deep learning models might be created that take into account financial news stories in addition to financial metrics like a closing price, traded volume, profit and loss statements, etc. for potentially improved outcomes.

II. Litreature Review

The first collection of articles include studies that primarily concentrate on stock market forecasting using artificial neural networks (ANNs). Biological neural networks serve as the basis for ANNs, which are computational models. Sets of nodes are organised into layers in the network, starting with an input layer and finishing with an output layer. As the connected nodes learn from instances and make an effort to lower the level of prediction error, signals are passed along (propagated) through them. Weights for the signals between connected nodes are changed as the system tries to increase its performance. The specific research focus and results of each ANN-related work are briefly described in the sections that follow.The unique research emphasis and results of the ANN study. Using data from multiple international stock markets, Jasic and Wood (2004) created an artificial neural network to forecast daily stock market index returns. The goal is to encourage lucrative trading. Untransformed data inputs are used in a system based on univariate neural networks that forecasts short-term stock market index return. The Standard and Poor's 500 Index (SP 500), the German DAX Index, the Japanese TOPIX index, and the London Financial Times Stock Exchange Index are all used in the study (FTSE All Share). The SP 500, DAX, and FTSE Index samples span the period from January 1, 1965 to November 11, 1999. The TOPIX sample spans the time from January 1, 1969, throughEnke and Thawornwong (2005) apply a machine learning information acquisition technique to assess the predicted correlations for various financial and economic variables.

variables. A ranking of the variables is created by calculating the information gain for each model variable. A cutoff is established to only include the most significant relevant variables in the forecasting models. neural network models

effective projection of future values. To increase the generalizability of various models, a cross-validation technique is also used. SP data from the 24-year period between March 1976 and December 1999 is used to compare the models. The findings demonstrate that, compared to the buy-and-hold strategy, alternative neural network models, and linear regression models, trading strategies governed by categorization models produce superior risk-adjusted gains.

III. Algorithums Used

A. *Multiple Linear Regression*

1. A more advanced variation of the simple linear regression process is multiple linear regression. The goal is to anticipate a variable's value based on the values of two or more other variables. The dependent variable's value must be predicted using the independent variables. Both continuous and categorical input independent variables are possible. The dependent variable is referred to as the variable for the prediction. When the independent variable is changed by one unit, the dependent variable also changes by the same amount.It is common practise to use this statistical technique to effectively forecast data, predict outcomes, analyse time series, and identify causal relationships between variables. It can be used to determine the impact of independent variables on a dependent variable. The prediction of a student's college GPA based on their CET score is one example of a practical application of multiple regression models. Multiple linear regression is a statistical method that has been around for a while and is utilised in stock market analysis.

2. Polynomial Regression: In statistics, a type of regression analysis known as a polynomial regression is one in which the relationship between the dependent variable y and the independent variable x is described as an nth degree polynomial in x. In order to get the additional predictors, each of the original predictors had to be raised in power. With an increasing value of n, more higher-order terms are present. Using it, users can fit a non-linear line to a set of data. Higher-order polynomials like square, cubic, quadratic, etc. are applied to one or more predictor variables to achieve this.In typical situations, one predictor and one outcome variable are employed. The relationship between the independent and dependent variables is known to be very well defined. Applications include the analysis of sediment isotopes, the investigation of health outcomes in medicine, and others. This dataset is non-linear in character and was used to train the model. Polynomial regression can be used to make predictions because, in the actual world, a stock market's growth is never linear like a line.

3. Random Forest Regressor: An algorithm for supervised learning is random forest. It is a regression and classification ensemble learning approach. An ensemble technique, as opposed to a single model, aggregates the predictions from various ML algorithms to provide predictions that are more accurate overall. Cross-validation allows for greater accuracy. The user can run as many trees as they want because random forests do not overfit.With increased dimensionality, users may precisely operate on a vast data set. Diabetes Prediction, Product Recommendation, Bitcoin Price Detection, Predicting Loan Defaults, and other prominent applications are a few examples. The accuracy is increased by the Meta estimator using the averaging technique while fitting several classifying decision trees on numerous dataset subsamples. A sequence

of the regression decision trees' results is combined to produce output forecasts. A mean of the forecasts made by the trees in the forest makes up an output prediction.

B. Flow in Proposed System

The steps in the flow of our proposed framework shown in the below Figure **??** have been explained in detail as follows: The suggested model for predicting daily stock values is based on the analysis of historical data, technical indicators, and optimising neural network algorithm. The suggested model architecture has six input vectors that reflect historical data and technical indications that were deduced, and one output vector that represents the upcoming price. The following procedures were used to create a neural network capable of forecasting short-term future actions of stocks, including their closing values.

(1) comprehension of the issue at hand and identification of crucial variables;

(2) pre-selection and collection of samples;

(3) input pre-processing;

(4) modelling and forecasting.

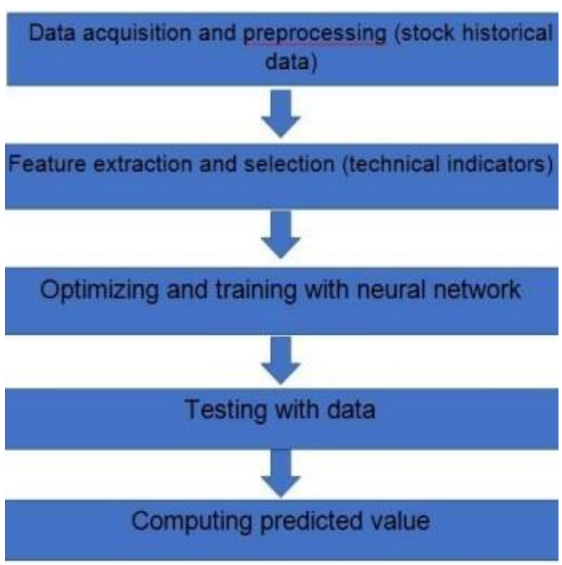


Fig. 1. Proposed System

Basic methods for creating and predicting values in Neural Works Predict using the Multi Layer Perception model. 1. Creating a Predict Model: To generate predictions from data where the desired outputs may fall within either a Continuous range of numeric values or a Discrete ordered range of numeric values. 2. Model selection: The Multi Layer Perception (MLP) model is chosen to forecast the stock price.

3. Training data for MLP input 4. Data from MLP Output Training 5. Features of MLP Training Data MLP Network characteristics 6. 7. Examining the parameters and developing the model Saving the model, step 8. 9. Information on training 10. Validation of a predictive model 11. Choosing the testing data sets. 12. Exam results interpretation Run an MLP predict model 13. 1. Missing Values removal- We should remove all the instances that have zero (0) as their value or contain any missing values as they are irrelevant and bring inconsistency to the input data. Therefore, these instances are eliminated. It also helps us to reduce dimensionality of data which helps us to decrease computational time in order to work faster.

2. Splitting of data- After cleaning the data, we generally normalize the data, i.e., we transform the columns of the input data set to the same scale to prepare the data for training and testing the machine learning model. After data splitting, we train the algorithm on the training data set and keep aside the test case data. This training process helps to correlate the processed output against the actual output. This helps us to modify the training model, which is the most suitable to achieve our goal.

1) Studies Using Hybrid: Some of the methods used to approach the problem of stock market prediction include ANNs, SVMs, and multi-method GA approaches. The works in this last category that have applied various distinct artificial intelligence approaches or several methods to this issue domain are described here.

2) Aritificial Intelligence: Although it is a crucial stock-related financial decision, asset allocation has gotten less attention in machine learning studies. A brand-new stock trading strategy is presented by O, Lee, Lee, and Zhang (2006) in a reinforcement-learning framework that also includes dynamic asset allocation. The suggested asset allocation approach, known as meta policy (MP), is built to make use of the temporal data from both stock recommendations and the stock fund's ratio over the asset. The design of an environment and learning agents facilitates the formulation of the MP in the reinforcement learning framework. The meta policy generator and every local trader are trained using data sets.

1. Analyze Stock Market: Many hybrid and neural network models have been presented in an effort to outperform conventional linear and nonlinear approaches for stock market forecasting, but there are some limits in most ANN model performance in this area.

A multi-layer perceptron (MLP), a dynamic artificial neural network (DAN2), and hybrid neural networks that use generalised autoregressive conditional heteroscedasticity (GARCH) to extract new input variables are all evaluated by Guresen, Kavakutlu, and Daim in their 2011 study on the effectiveness of these techniques. NASDAQ index readings for every day between October 7, 2008, and June

26, 2009, are used to compare various techniques. One finding is that it appears that the simple MLP is the best and most useful ANN design. A method to forecast the daily return direction of a group of stocks is presented by Zhong and Enke (2019). To forecast the daily direction of future stock market index returns, deep neural networks (DNNs), conventional ANNs, and two datasets altered using principal component analysis (PCA) are all applied to the whole preprocessed but untransformed dataset. As the number of hidden layers gradually grows from 12 to 1000, a pattern for the classification accuracy of the DNNs is found and shown while controlling for overfitting. Simulation results demonstrate that compared to other hybrid machine learning algorithms or the complete untransformed dataset, DNNs using two PCA-represented datasets provide much higher classification accuracy.

Steps for Decision Tree Algorithm:-

Step-1: Start creating the tree with the root node, say S, which contains the input dataset.
Step-2: Use Attribute Selection Measure (ASM) like gini index or information gain, etc. to find the best attribute in the dataset
Step-3: Divide the S into subsets based on all possible values for the best attribute obtained via the decision taken in step-2.
Step-4: Generate the decision sub-tree using the best attribute node.
Step-5: Recursively generate new decision sub-trees using the subsets of value attributes of the data set created in step-3. Continue this process until we cannot further classify the nodes i.e. we reach a leaf node.

4. Random Forest: Random Forest Algorithm mainly is based on the concept of ensemble learning in which we combine multiple classifiers with each other in order to solve a complex problem which improves the performance of the model. Random Forest is a combination of multiple number of decision trees applied on various subsets of the input data set which then takes their average, i.e., it uses the prediction from each decision tree and based on the predictions which get the majority of votes, it predicts the final output.

5. Support Vector Machine (SVM): SVM creates a hyperplane or set of hyperplanes in a N-dimensional space, which separates two classes and classifies the data points into those classes. There can be multiple decision boundaries to segregate the classes in the N-dimensional space, but we only find out the best decision boundary called the hyperplane of the SVM to classify the data points. The dimensions of the hyperplane depends on the number of input features present in the considered data set. So for e.g. if there are 2 input features, then the hyperplane will be in the form of a straight line. And if there are 3 input features, then the resultant hyperplane will be a 2-dimensional plane. Our goal is to always create a hyperplane such that we get the maximum margin

3) *AI Model:* Describe the layout and structure of a trading signal mining platform that uses an extreme learning machine (ELM) to determine stock prices.

forecasts made from two sources of data simultaneously. On the basis of intra-day data from the H-share market (shares of mainland Chinese companies listed on the Hong Kong Stock Exchange) and recent news archives, experimental comparisons between ELM, SVMs, and backpropagation neural networks (BPNNs) are conducted.The findings demonstrate that compared to BPNN, RBF ELM and RBF SVM both obtain higher prediction accuracy and faster prediction speed. Additionally, RBF ELM outperforms RBF SVM in terms of prediction speed. Data from two sources—a Caihua market news archive and H-share market stock prices from 2022—is used to simulate a basic trading strategy with the signals.

IV. RESULTS

exemplifies an artificial neural network's example of a resolution-appropriate image. In order for connected information processing units to convert input into output, they need a model known as a neural network that has an activation function. Raw input is taken in by the first layer of the neural network, which then processes it and sends the results to the hidden layer.

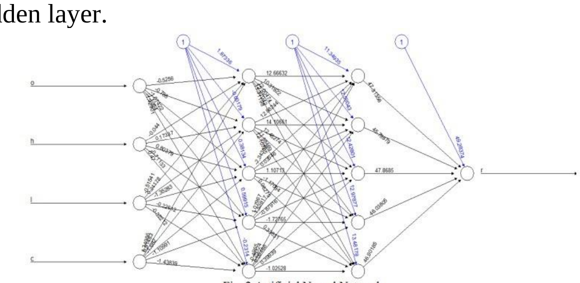
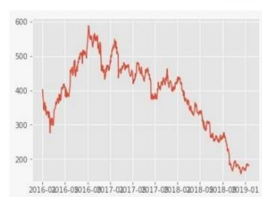


Fig. 2.Artificial Neural Network



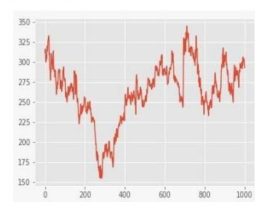


Image title.

The prediction graph employing a neural network is shown in Figs. and 3 the opening price from January 2020 to January 2021 is shown in Fig. of the Tata Motors data. The closing price of Tata Motors is also shown in fig. . The data was retrieved from the Yahoo server.

V. CONCLUSION AND FUTURE SCOPE

A survey of recent literature will be used to determine future directions for machine learning stock market prediction research. There are various inferences that may be drawn about our current understanding of this field of study given the ML-related systems, problem contexts, and discoveries reported in each chosen article as well as the taxonomy categories that were previously presented. First, there is a clear connection between ML techniques and the prediction issues they are used to solve. Similar to task-technology fit (Goodhue and Thompson, 1995), where the right match between tasks and For anticipating numerical stock market index values, artificial neural networks work well. Determine if the total stock market index is expected to climb or fall using support vector machines, for example. In order to find solutions to problems, genetic algorithms use an evolutionary method.higher quality system inputs, or forecast which stocks to include in a portfolio, to achieve the best results. Despite the fact that each study showed how the strategies might be used successfully, single method applications do have some restrictions. One approach to alleviate some of these restrictions is to use hybrid machine learning approaches. The issue is that eventually the systems get so complicated that they are useless in real-world applications. Future research can address this theoretical and practical issue.The second finding from this evaluation of prior research is the need for better generalizability of results. Most studies just look at one market and/or one time period when evaluating an ML system, not taking into account how well the system will work in other circumstances. The experimental system evaluation can be improved in three ways. First off, the outcomes from Asian stock markets serve as the foundation for many studies. In the same time frame, these technologies might likewise be tested for the US or European markets. Second, the systems' performance in various market settings might be evaluated using data from periods when markets are increasing or falling.Would a method be able to effectively forecast market values in the US both during the recent market expansion from 2018–2019 and the financial crisis of 2008–2009, for instance? Do systems also have the ability to foresee market decline if they can forecast market growth? Last but not least, the suggested methodologies might be used to compare the forecasting performance of stock market indices that only include large corporations vs only small enterprises. System performance in various risk and volatility environments A stronger research and practise contribution will result from any of these improvements to the experimental approach.Reflection also made it clear what the ultimate findings were. The inputs, algorithms, and performance metrics of ML systems require a more robust underpinning of financial investment theory. In the absence of this, the outcomes can just be.

References

Cabitza, F., Locoro, A., Banfi, G. (2020). Machine learning in orthopedics: a literature review. Frontiers in Bioengineering and Biotechnology, 6, 75. Chavan, P. S., Patil, S. T. (2019). Parameters for stock market prediction. International Journal of Computer Technology and Applications, 4(2), 337. Chiu, D. Y., Chen, P. J. (2020). Dynamically exploring internal mechanism of stock market by fuzzy-based support vector machines with high dimension input space and genetic algorithm. Expert Systems with Applications, 36(2), 1240-1248. Chong, E., Han, C., Park, F. C. (2019). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83, 187-205. Dai, W., Wu, J. Y., Lu, *C.* J. (2021). Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. Expert systems with applications, 39(4), 4444-4452. Das, S. P., Padhy, S. (2021). Support vector machines for prediction of futures prices in Indian stock market. International Journal of Computer Applications, 41(3).

Dash, R., Dash, P. K. (2021). A hybrid stock trading framework integrating technical analysis with machine learning techniques. The Journal of Finance and Data Science, 2(1), 42-57. Enke, D., Thawornwong, S. (2021). The use of data mining and neural networks for forecasting stock market returns. Expert Systems with applications, 29(4), 927-940. Goodhue, D. L., Thompson, R. L. (2000). Task-technology fit and individual performance. MIS quarterly, 213-236. Guresen, E., Kayakutlu, G., Daim, T. U. (2021). Using artificial neural network models in stock market index prediction. Expert Systems with Applications, 38(8), 10389-10397.