DOI:10.15613/hijrh/2022/v9i2/218195
ISSN (Print): 2349-4778
HuSS: International Journal of Research in Humanities and Social Sciences, *Vol 9(2), 89-94, July-December 2022* ISSN (Online): 2349-8900

Machine Learning and Deep Learning based Approaches to Predict Nifty Index

P. Karthikeyan[1]* and N. Vigneshwaran[2]

[1]Associate Professor, Department of Management Studies, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India; ptp_karthi.mba@kongu.edu
[2]II MBA, Department of Management Studies, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India; vigneswarann.21mba@kongu.edu

Abstract

Stock price prediction is one of the most difficult machine learning issues to solve. The stock market, often known as the equity market, has a significant impact on today's economy. This study discusses about different machine learning and deep learning approaches to predict and evaluate stock prices. Time series data is used to depict stock values and algorithms are trained to learn patterns from trends. For the machine learning approaches, study used linear regression, logistic regression and decision tree and for deep learning approaches Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN) are used to predict Nifty Index value. These variants are commonly used to forecast stock prices and movements. The algorithm is based on the concept of probability and it is used for predictive analysis. For continuous quantitative data, a regression tree is utilized. Linear Regression, Logistic Regression, Decision Tree, LSTM and RNN are the most noticeable techniques used in financial time series forecasting. The study observed from Python software, that the Linear and Logistic Regression model predicts accuracy of roughly 52% and provides an acceptable return ratio. As a result, study found that the Nifty-50 data set has been utilized to improve the precision of supervised learning and future prediction.

Keywords: Decision Tree, Deep Learning, Linear Regression, Logistic Regression and Nifty

1. Introduction

A share market is a market where publicly traded firms or individuals' shares are traded to raise funds. Stock market forecasting is the process of attempting to forecast a company's future stock value. Stock market index forecasting is a significant financial issue that is garnering more attention[6]. A multitude of methods, such as linear and logistic regression are used by statistical researchers to predict average numerical models[15]. The value of a class or group of persons is predicted using logistic regression[8]. For classification of the data set, the regression and decision tree of non-parametric supervised learning method are used[2].

The goal is to use simple decision rules abstracted from data structures to improve a model that accurately predicts the value of a target variable[3]. Stock prices and index rates are difficult to predict and they are influenced by a variety of factors such as political situations, the global economy, financial reports and company performance and so on. The many tactics for projecting stock prices are analyzed and the proper patterns are established based on stock market movements, in order to maximise profits and avoid losses[4]. Using closing and opening prices, volume traded, adjacent close values and other technical analysis methods, the stock market return is predicted. It is also used to predict the cost of a stock in the future. When compared to prior methods, machine learning techniques have proved to increase efficiency by 60-86%[2].

Linear Regression, Logistic Regression, Decision Tree[2], RNN and LSTM[11], techniques are used in various

*Author for correspondence

financial data and to create new factors that are used as inputs of the model[9]. For processing very large data sets at very fast time, such systems often analyze individual stock data as a time series rather than studying correlations and patterns between different businesses. The study has developed a model for investing in the stock market that doesn't require any previous trading expertise[3].

Based on previous prices, the model predicts the stock index's trend. The four strategies namely Technical Evaluation, Analytical Framework, Analysis of Time Series and Methods of Machine Learning are used to anticipate the stock market in general. Technical analysis is a method of detecting trends by plotting technical data such as Open, High, Low and Close prices on charts[2]. As a result, each expert examines the same chart and offers a distinct trading perspective. Machine learning and algorithms are used to make predictions about market movements[12].

2. Review of Literature

Vatsal H. Sham[12] analyzed and forecasted the stock market values by examining the past years' share price movement and pattern. This would help the individuals to know about the ways to increase profit by reducing losses. Sirignano, J[11] analyzed the modeling in a time series using nonlinear parameter fitting process. The LSTM model was a nonlinear time-series model used for effective decision making of the stock market prediction. The LSTM model, according to Rajashree Dash[6] played a crucial role in RNN versions. The major goal of this model was to employ self-loop design to create an incline road that might be used constantly throughout time. Every iteration would update the self-weight, loops preventing the gradient vanishing issue.

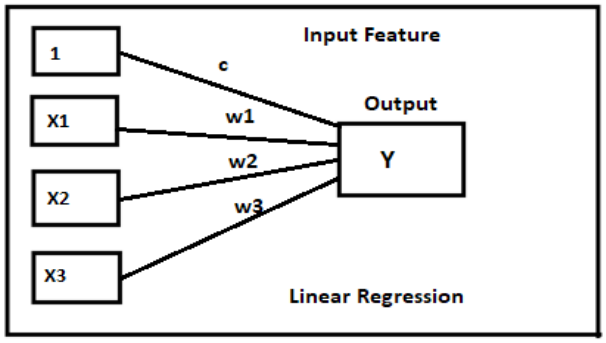
Deep learning, according to Dutta and Shekhar[7], was a highly complicated nonlinear relationship that closely described manipulating components in complex attributes. The accuracy of the deep learning model had been tested across a wide range of fields. Zhang and Zhou[5] had mostly concentrated on NIFTY trend prediction (upward/downward). Vatsal H. Sham[12] had included Strike Price, Last Traded Price (LTP), Implied Volatility (IV), Open Interest (OI) and Traded Volume for predicting the market price of the respective company's shares. Dutta and Shekhar[7] had studied stock market trends from huge volumes of dataset and it represented the fundamental stock price movement. Recurrent Neural Network (RNN) algorithms were widely used and played a vital part in stock market index prediction decision-making. Sreelekshmy[10] and Zhang[5] had analyzed the numerical analysis of the stock trend, but also the textual analysis of it by assessing public opinion from online news sources and blogs. Ajit Kumar[6] analyzed the time series data to represent stock values and Neural Networks were used to learn patterns from trends. Sirignano[11] had applied self-learning process, deep learning algorithms for finding hidden patterns and underlying dynamics in data. Sreelekshmy[10] had studied RNN and LSTM architectures for detecting long-term dependencies and used them to forecast the future. Haines et al.,[8] had applied window strategy to forecast future values over the short term. The models' performance was measured using percentage error.

3. Proposed System

3.1 Linear Regression

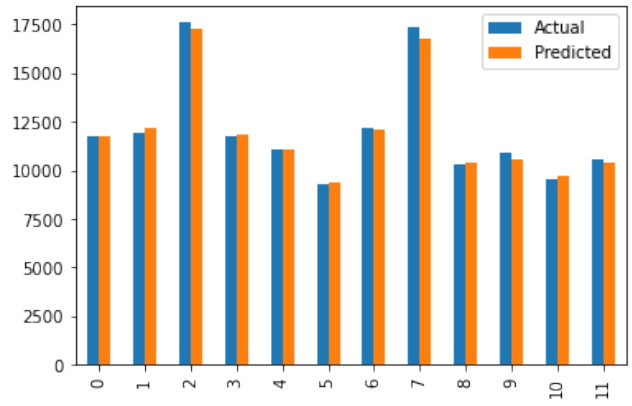
By fitting a linear equation to observed data, linear regression seeks to model the relationship between two variables. One variable is used to explain the outcome, while the other is used to determine it[2]. Linear regression is a supervised machine learning technique that generates a projected output with a constant slope. Figure 1 shows the linear regression of the data set.

It is used to forecast values within a continuous range (e.g. Open, High, Low, and Close prices) rather than categorizing them (e.g. Technical Analysis, Fundamental



Source : Computed

Figure 1. Linear Regression.



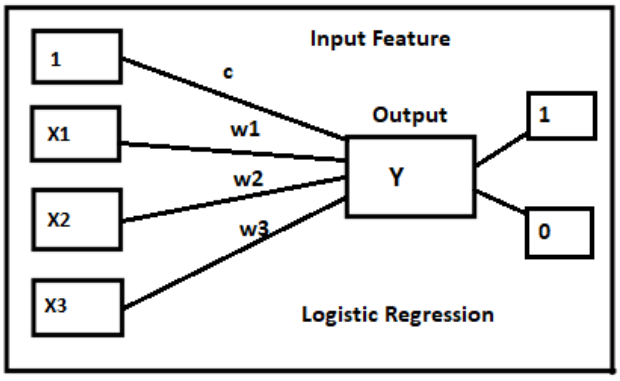
Source : Computed

Figure 2. A comparison and contrast of actual and predicted values.

Analysis, Time Series Analysis)[2] because it investigates two independent variables to find a single relationship. Linear regression is a useful tool for technical and quantitative analysis in financial markets. A trader can utilize linear regression to find important price points like entry, stop loss, and exit[4]. Figure 2 gives the pictorial representation of actual and predicted values.

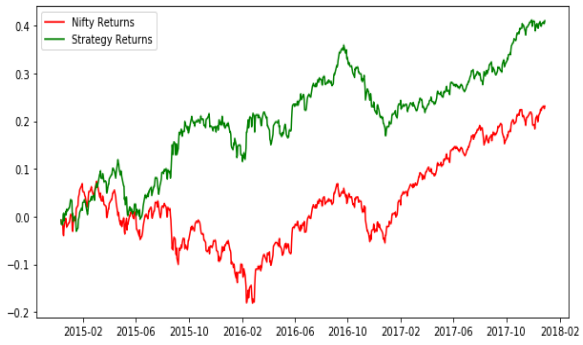
3.2 Logistic Regression

Logistic Regression is a categorization problem-solving Machine Learning method. It is a probability-based predictive analytic method. According to the logistic regression hypothesis, the cost function should be confined to a value between 0 and 1[8]. Figure 3 shows the logistic regression of the data set.



Source : Computed

Figure 3. Logistic Regression.



Source : Computed

Figure 4. Stock price return.

	Predicted: -1	Predicted: 1
Actual: -1	103	251
Actual: 1	101	290

Source : Computed

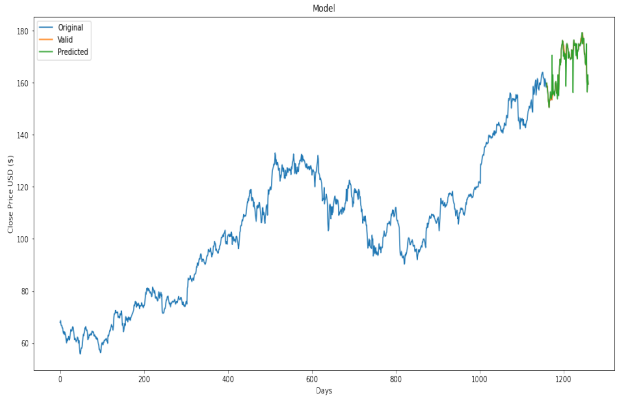
Figure 5. Actual and Predicted Values

Stock marketplace motion can be anticipated with the use of logistic regression, which could forecast growing trend, an unchanged trend or a falling trend. The class of supervised studying consists of logistic regression. The problem of predicting the motion of inventory prices[8] is given in Figure 4.

The trader will buy (1) if the stock's tomorrow's closing price is greater than today's closing price; else, trader will sell it (-1). If the outcome is 0.52, the study says that there is a 52% chance that tomorrow's closing price will be greater than today's closing price, and the study will give 1 rating (high probability)[3,4] as given in Figure 5.

3.3 Decision Tree

It is reasonably simple to learn and is highly successful. It is frequently used in predictive modeling[15]. A regression tree is used to forecast continuous quantitative data. Accuracy can be calculated by comparing actual test set values to expected values. So, 67.53% categorization rate, is considered to be fairly accurate. The parameters of the Decision Tree Algorithm can be modified to increase accuracy[2] as shown in Figure 6.



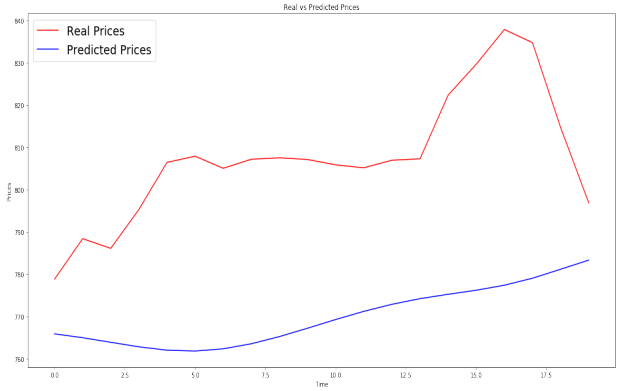
Source : Computed

Figure 6. Classification of two step process

3.4 Long Short-Term Memory (LSTM)

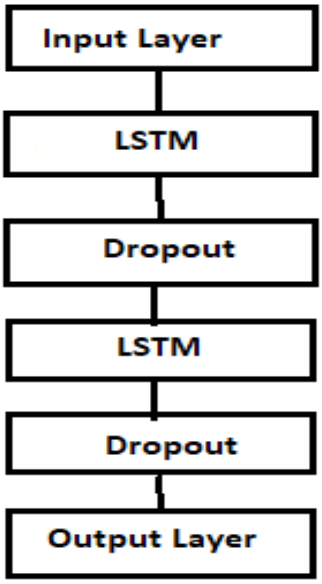
A regression tree is used to forecast continuous quantitative data. Accuracy can be calculated by comparing actual test set values to expected values. So, 67.53% categorization rate is considered to be fairly accurate. The parameters of the Decision Tree Algorithm can be modified to increase accuracy[10]. The scale of the data is very important to LSTM and the scale of the Close value is in a kind of scale, so the researcher should always try to modify the value. The values will be transformed from 0 to 1 using the min-max scalar[6].

Researcher has to restructure so that fit transform can be used. It should partition the data differently when



Source : Computed

Figure 7. Classification of predicting real and future price value.



Source : Computed

Figure 8. The Process of initializing the LSTM model.

training time series data and the data using appropriate date. In time-series data, one variable is dependent on another[3]. The test size should be the difference between the length of the dataset and the training size; the training size should be 65% of the overall length of the data frame. If researcher wants to estimate the price of a stock in a day, then researcher need to think about the time steps and how the preceeding data should be used. The value of the time step will now be set to 100. The information is classified into two categories X and Y as two different types of people[11].

The first 100 elements are used as the initial record in the 0^{th} iteration, while the 101^{st} element is placed in the X. The Y will be used to display the 100 components. X has to be always restructured in 3-D and add 1 when implementing any LSTM[6]. The reason for this is the time step and 1 is supplied to the LSTM. The input gate adds fresh information to the cell state. It gets rid of data that the model no longer requires. The Output Gate of the LSTM selects the data to be displayed as output[4] which is shown in Figure 7.

The initialization of the model is followed by the first input layer and the LSTM layer with a 0.2% dropout. The second LSTM layer has a dropout of 0.2%. With a 0.2% dropout, the third LSTM layer is used. The study uses return sequence true in the last layers since study does

not want to use the prior output. Fourth LSTM layer with 0.2% dropout study will not pass any activation to the output layer because it uses a continuous value model. The network is being built. The calculation of the best parameters and fitting the model[4] is given in Figure 8.

3.5 Recurrent Neural Network (RNN)

When the model uses RNN, Long Short-Term Memory (LSTM), Machine Learning and Deep Learning models, the stock price predictions will be more accurate[10].

A RNN is a form of advanced neural network with internal memory and the ability to analyze long sequences[13]. This makes RNN ideal for stock price prediction, which requires a large amount of previous data. Figure 9 depicts the comparison of predicted and expected value using simple RNN and LSTM models.

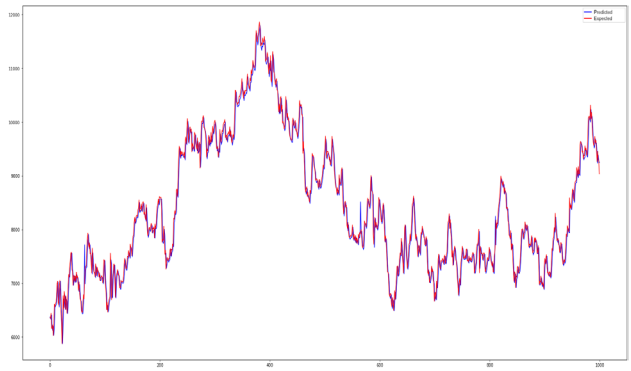
4. Experiments and Results

4.1 Dataset

The Nifty-50[14] data collection consists of highly traded stocks from three different industries. Each includes the information namely stock symbol, stock series, stock date and previous closing, opening, high, low, last, closing and average prices, total traded amount, turnover and number of trades. The day-by-day closing price of each stock is exclusively extracted from these datasets because day-by-day stock prices are preferred by the investors based on their decisions on the market's closing price[9].

The data is then standardized to bring the range of values closer to 0 and 1. Data is normalised to bring all stock data into a single range as the stock data from several market places is used and hence, the data must be within a specified range[10].

The study tries to predict the model using Nifty-50 stock data[14] to test if the models could detect common stock exchange movements. Yahoo Finance[8] provides the information. In the New York Stock Exchange, the top two most active stocks are chosen. Data from the current year has been used to construct the dataset. Only the day-by-day closing price is collected. For the current years, day-by-day closing prices for each company are used. The extracted data is then standardized using an algorithm



Source : Computed

Figure 9. Comparing the predicted value and expected value.

to unify the data before being sent into the network[6,10] . Figure 10 gives an overview on the code.

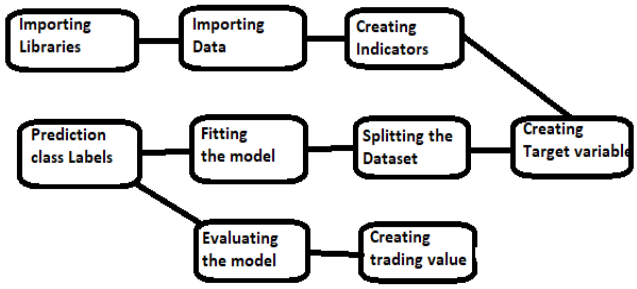


Figure 10. An Overview on Coding.

5. Conclusion

The dataset has been used to implement Linear Regression, Logistic Regression, Decision Tree, LSTM and Recurrent Neural Network (RNN)[8,15]. As expected, the future value calculations have been examined. The accuracy rate is defined as the number of accurate classifications divided by the total number of classifications[3].

The term "training accuracy" refers to how well the model performs when given historical data as an example. The performance of the stock value with unknown input is known as testing accuracy. The following metrics of confusion matrix can be used to calculate accuracy. It's a method for calculating a classification's work efficiency in stock prediction[4].

Prediction of the formula described in equation (1) gives the value for Linear Regression:

$$Y = b X + a \quad \ldots\ldots\ldots \quad (1)$$

Prediction of the formula indicated in equation (2) gives the value for Logistic Regression:

$$\mathbf{Y = 1 / 1 + e^{-x}} \quad \ldots\ldots\ldots (2)$$

6. References

1. Kaya MIY, Karsligil ME. 2nd IEEE International Conference on Information and Financial Engineering (ICIFE). 2010; 478–482.
2. Tsai, Wang S. Stock Price Forecasting by Hybrid Machine Learning Techniques. Proceedings of the International Multi Conference of Engineers and Computer Scientists. 2009.
3. Zhang X, Fuehres H, Gloor PA. Predicting Stock Market Indicators through Twitter "I hope it is not as bad as I fear". Procedia-Social and Behavioral Sciences. 2011; 55-62. https://doi.org/10.1016/j.sbspro.2011.10.562
4. Nikfarjam A, Emadzadeh E, Muthaiyah S. Text Mining Approaches for Stock Market Prediction. The 2nd International Conference on Computer and Automation Engineering (ICCAE). 2010; 4:256–260. https://doi.org/10.1109/ICCAE.2010.5451705
5. Zhang D, Zhou L. Discovering Golden Nuggets: Data mining in Financial Application, Systems, Man and Cybernetics, Part C: Applications and Reviews. IEEE Transactions. 2004; 34(4):513–522. https://doi.org/10.1109/TSMCC.2004.829279
6. Rout AK, Dash PK, Rajashree D, Bisoi R. Forecasting Financial Time Series using a Low Complexity Recurrent Neural Network and Evolutionary Learning Approach. Journal of King Saud University-Computer and Information Sciences.2017; 29 (4):536–552. https://doi.org/10.1016/j.jksuci.2015.06.002
7. Dutta S, Shekhar S. Bond rating: A Non-conservative Application of Neural Networks. Proceedings of the IEEE International Conference on Neural Networks. 1998; 2:443–450.
8. Haines LM, et.al. D-optimal Designs for Logistic Regression in Two Variables, mODa 8- Advances in Model-Oriented Designed and Anaysis. Physica-Verlag HD. 2007; 91–98. https://doi.org/10.1007/978-3-7908-1952-6_12
9. Yahoo Finance - Business Finance Stock Market News, [Accessed on August 16, 2018].
10. Selvin S, Vinayakumar R, Gopalakrishnan EA, Menon VK, Soman KP. Stock Price Prediction using LSTM, RNN Mode. International Conference on Advances in Computing, Communications and Informatics (ICACCI). 2017; 1643–1647. https://doi.org/10.1109/ICACCI.2017.8126078
11. Sirignano J, Cont R. Universal features of Features of Price Formation in Financial Markets: Perspectives from Deep Learning. Quant. Financ. 2019; 19:1449–1459. https://doi.org/10.1080/14697688.2019.1622295
12. Sham VH. Machine Learning Technique for Stock Market Prediction.
13. Khan ZH, Alin TS, Md. Hussain MA. Price Prediction of share market using Artificial Neural Network. International Journal of Computer Application. 2011; 22(2): 42–47. https://doi.org/10.5120/2552-3497
14. Abinaya P, Kumar VS, Balasubramanian P, Menon VK. Measuring Stock Price and Trading Volume Causality among Nifty 50 Stocks: The Toda Yamamoto Method. International Conference on Advances in Computing, Communications and Informatics (ICACCI). 2016:1886–1890. https://doi.org/10.1109/ICACCI.2016.7732325
15. Li L, Wu Y, Ou Y, Li Q, Zhou Y, Chen D. Research on Machine Learning Algorithms and Feature Extraction for Time Series. IEEE 28th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). 2018:1–5. https://doi.org/10.1109/PIMRC.2017.8292668