Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

Machine Learning Algorithm for Stock Market Prediction – A Comparison

A. Gokul[1], R. Manoj Kumar[2], A. Jayasurya[3]

[1,2]Assistant Professor, Department of Computer Science and Engineering, Paavai College of Engineering, Pachal, Namakkal

[3]Assistant Professor, Department of Computer Science, Excel College for Commerce and Science, Komarapalayam, Namakkal

Email: [1]gokuldominator@gmail.com

Abstract

The prediction about the stock market serves as an effort to forecast the value of the market, an individual stock, or a particular industrial sector. The prediction or forecast is usually done by using several approaches and analyzing the fundamental or technical details of an industry, an economy, or both. Predicting stock markets is very essential, as successful prediction can help in proper decision-making as well as in increasing the profit of the business. As prediction of the stock market is a bit complicated and challenging, conventional methods do not consistently forecast the changes with absolute certainty. For this reason, the proposed study has come up with a comparison of the machine learning models for forecasting the market changes. The aim of the study is to compare two machine learning models: the decision tree (DT) algorithm, which excels at handling nonlinear relationships and feature interaction, and the long short-term memory (LSTM) algorithm, which could efficiently capture long-range dependencies in sequential data. The Yahoo Finance dataset was used for performing the comparison. The results observed will be utilized to analyze the stock expenses and their predictions in the future.

Keywords: Stock Exchange, Yahoo Finance, DT, LSTM, Machine Learning Accuracy.

DOI: https://doi.org/10.36548/jaicn.2023.4.005
Received: 17.10.2023, received in revised form: 22.11.2023, accepted: 01.12.2023, published: 13.12.2023

1. Introduction

Predicting the stock market value of a particular industry, an individual stock, or any other asset that is traded on a market to have knowledge about its future value has become very essential nowadays in business to make proper investments and gain expected returns. Successful forecasting and prediction always lead to significant profit. According to the efficient-market theory, stock prices are a reflection of all information that is currently available, and any price fluctuations that are not based on recently disclosed information are therefore, by their very nature, unpredictable. There are other numerous factors, such as economic statistics, geopolitical developments, investor mental states, and unanticipated events, that can affect the stock market. The underlying noisy conditions and high fluctuations with regard to market trends also impede stock market prediction analysis. So, in the rapidly evolving industrial world, accurately predicting stock trends is both an intriguing and challenging task. Several factors, both economic and non-economic, are taken into account and have an impact on how stock movements behave. So, predicting the stock market becomes more challenging while taking into consideration production and profit increases [8]. There several approaches used in predicting the movements of the stock market. They are as follows. The table.1 below shows the different types of approaches commonly used in predicting the stock market.

Table 1. Prediction Methods

Prediction Methods	**Uses**
Fundamental Analysis	Attempts to determine stock's true value that is compared to the value at which it is traded on stock exchanges to determine whether or not the stock is cheap. To ascertain a company's intrinsic value, this entails looking at its earnings, dividends, financial stability, and general market circumstances. It is also referred as the top-down analysis. It is a long-term strategy.
Technical Analysis	It is not concerned about the fundamental details of the organization Relies on the past price charts and the volume of trade to identify the patterns and the trends.

Technological Analysis	This involved the machine learning approaches (uses both the fundamental and the technical details) Utilizes large dataset about the stock market, financial time series etc to learn the pattern that indicate the future stock movements.

In recent days many businesses and the industrial sector has started using the machine learning techniques in forecasting their stock value and the fluctuations in the market as machine learning offers significant potential in forecasting the stock market and can be easily integrated into the business intelligence as well as help in real-life decision making. Leung, et al [11] In his work, used SSVM to forecast whether the stock prices of the inputs will move in a positive or negative direction. The complicated inputs like the nodes of a graph structure are classified using structural support vector machines (SSVMs), Box GEP et al [1] offers techniques for creating, determining, fitting, and verifying models for time series and dynamic systems in which system observation and the chance to exercise control occur at regularly spaced time intervals. Huang et al [2] proposed a new wavelet kernel based SVM to evaluate their predicting abilities, and forecast the Nasdaq composite index. Wu,et al [3] on his survey conducted with experiment over the "Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) dataset collected" and concluded that the hybrid computational intelligence methods can be highly recommended for the stock price forecasting as the methods have resulted with low MAD, MAPE and RMSE. Kao, et al [4] shows that his Wavelet-MARS-SVR stock prices forecasting model is capable of solving the problem of wavelet subseries and offers efficient performance. Soni, et al [5] explores the various methods used to anticipate share values, ranging from neural networks to graph-based approaches to traditional machine learning and deep learning techniques. It provides a thorough review of the methods used to forecast stock values and looks at the difficulties involved as well as the direction that the field may take in the future. Mintarya, et al [6] in his study provided the review of relevant literature regarding machine learning techniques (SVM and Neural networks) for stock market prediction. Patel, et al [7] discussed about variety of machine learning techniques, both supervised and unsupervised, and how investors can learn about changes in stock prices and how to calculate accuracy.

The objective of the proposed study is to compare the ability of two machine learning algorithm LSTM and the DT in forecasting the stock market utilizing the yahoo finance dataset (2018 -2023) that was collected from Kaggle [12]. These algorithms DT and LSTM

were preferred in the study among other because the decision tree (DT) algorithm, excels at handling nonlinear relationships and feature interaction, and the long short-term memory (LSTM) algorithm, could efficiently capture long-range dependencies in sequential data.

1.1 Technical Objective

The models are implemented using the programming language Scala, the dataset details include the expenses and the cost of the shares in the past, and the market index to estimate the cost of the shares.

1.2 Dataset Description

The Yahoo Finance dataset (2018 -2023) from Kaggle was used in the study to compare the two models DT and LSTM. This "yahoo_finance_dataset(2018–2023)" includes daily stock market data for a variety of assets, including stocks, exchange-traded funds, and indexes. Its dimensions are 1257 rows by 7 columns, and its duration is April 1, 2018, through March 31, 2023. The dataset, which was obtained from Yahoo Finance, aims to give academics, analysts, and investors access to a comprehensive dataset for the analysis of stock market trends, pattern recognition, and the creation of investing strategies. Because the dataset is supplied in XLSX format, importing it into different data analysis programs is simple. The fundamental data points like open, close, high, low, adjusted close and the volume are obtained from the dataset directly. These fundamental data points are used in the analyzing the individual stock movements. The technical indices are derived from these data points to analyze general market trends and possible trading situations. Moving averages, the Relative Strength Index (RSI), the Moving Average Convergence Divergence (MACD), and other indices are examples of common technical indices. To find patterns and trends in the market, these indexes are computed using historical price and volume data. They provide information on market strength, momentum, and possible buy or sell signals. Technical indices provide a more comprehensive understanding of market movements, even when the fundamental data points pertain only to particular stocks.

The input data structure is generated by calculating the related values and the indicator values using the mathematical formulas of the technical indices Further the dataset is divided for the purpose of training and testing, three forth of the dataset was used in training and one-

fourth was used in testing. The figure.1 below shows the overall process involved in the proposed study.

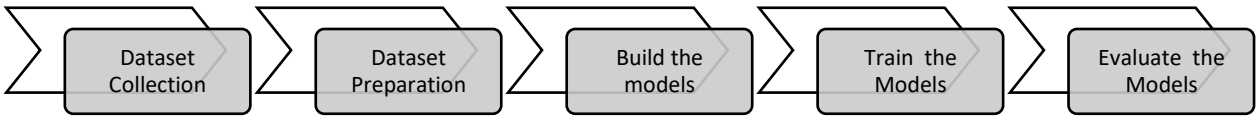

Figure 1. Overall Process of the Study

2. Model Building

The two machine learning models LSTM and DT are considered in the study, along with the technical indices like Moving Averages, Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Bollinger Bands, and many others related to stock market. The technical indicators mentioned above are used as input variables for the Machine learning models

2.1 Decision Tree

The Decision Tree model is trained on past financial data, which includes prices and also the technical indicator values. The DT learns from the historical data's relationships and patterns. The Decision Tree can be used for prediction or decision-making once it has been trained. It could be used, for instance, to forecast whether the price of a financial instrument would rise, fall, or stay the same based on the present values of technical indicators. The interpretability of decision trees is one of its benefits. The decision rules produced by the tree can be interpreted by traders and analysts to determine which combinations of technical indicator values are connected to particular market circumstances or price movements. It also capable of determining the entry and the exit points of the trades. The simple decision tree model is used by importing the necessaries libraries in scala.

2.1.1 Decision Tree in Scala

The following steps provides the outline of the implementation of DT in Scala for the stock market prediction using the yahoo finance dataset (2018 -2023) collected for the last 60 months.

Step 1: Import the necessary libraires, the study uses the Apache Spark library for Decision tree

Step 2: Set the parameters, Maximum_Depth and the Minimum_Number_ of _Samples.

Step2: Load the Dataset collected from yahoo finance in XLSX format. [date, open, fall, close, adjusted close, volume]

Step 3: Prepare the data by removing the outliers and imputing the missing values.

Step 4: Extract the relevant features, convert the variables to numerical representation using one hot encoding and scale the features.

Step 5: Split the dataset, the total number of datasets is split as 75 % for training and 25% for testing.

Step 6: Train the decision tree model using the training dataset.

Step 7: Deploy the model for the purpose of testing.

2.2 Long Short-Term Memory

Long Short-Term Memory (LSTM) are a type of Recurrent neural networks (RNNs) useful for time series forecasting, which includes stock market prediction. RNNs are well-suited for modeling sequential data. LSTMs are capable of capturing intricate patterns and dependencies in financial data when paired with technical indicators. The generated input data structure is used as input in the LSTM model, along with the technical indicators including the RSI, MACD, and Bollinger Bands. The data are normalized using the Scala and converted into a sequence to make it suitable for training. Every sequence holds the window of the past data points and its corresponding technical indicator values.

2.2.1 LSTM in Scala

The following steps provides the outline of the implementation of LSTM in Scala for the stock market prediction using the yahoo finance dataset (2018 -2023) collected for the last 60 months.

Step 1: Import the necessary libraires and its dependencies, the study uses the Deeplearning4j. and its dependencies for LSTM, the built tool SBT was used to manage the dependencies

Step2: Load the Dataset collected from yahoo finance dataset in XLSX format. [date, open, fall, close, volume]

Step 3: Split the dataset, the total number of datasets is split as 75 % for training and 25% for testing.

Step 4: Normalize the dataset using the Min-Max scaling.

Step 5: Convert the dataset into sequence such that Every sequence holds the window of the past data points and its corresponding technical indicator values.

Step 6: Build and Train the LSTM using the training dataset. (Tanh function is used in determining the new cell state and sigmoid to determine the hidden state)

Step 7 : Deploy the model for the purpose of testing.

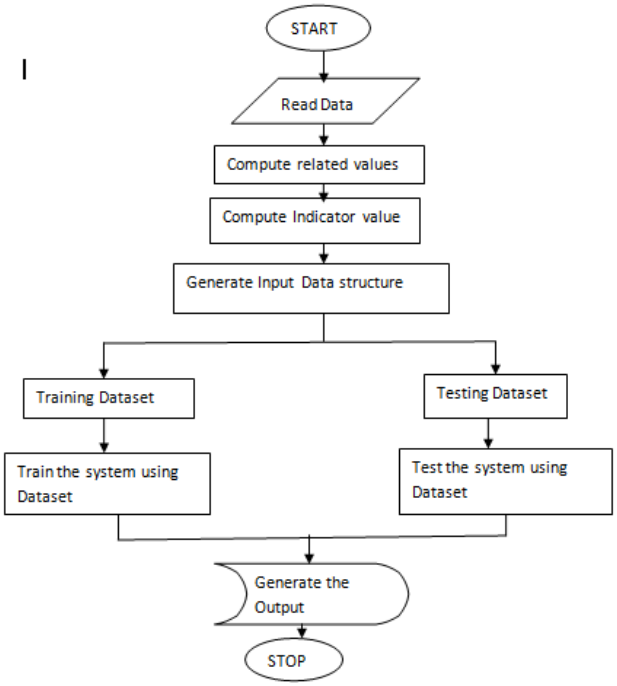

Figure 2. Flow Diagram

The Scala uses different libraries Apache Spark and Deeplearning4j for DT and LSTM respectively. The general flow diagram of the stock market prediction is shown in Figure.2

3. Technical Indices

The section presents the most commonly used technical indices of the stock market prediction, some of the technical indices that are used in the study is as follows. The detail of the technical indices used are as follows.

3.1 Moving Averages (MA)

Statistical computations called moving averages are used to examine data points over a given amount of time. They are frequently used in the stock market to even out price movements and pinpoint possible entry or departure opportunities by looking at the average price over a given period of time.

Simple moving average is calculated using the $SMA(n)\frac{x1,x2,x3\ldots xn}{n}$, where SMA(n) is moving average for n periods and the x1, x2 are data points of each period

3.2 Relative Strength Index (RSI)

The momentum oscillator, or RSI, measures the rate and direction of price changes. It is employed to determine whether a market is overbought or oversold. RSI is frequently used by traders to identify possible trend reversals.

The Relative Strength Index (RSI) is typically determined utilizing the mathematical expression:

RSI=100- 100/1+RS

RS= Average Gain/ Average Loss

Average Gain =Sum of gains over the last n periods/n

Average Loss =Sum of losses over the last n periods/n

3.3 Moving Average Convergence Divergence (MACD)

MACD is a momentum indicator that follows trends and displays the relationship between two moving averages of the price of an asset. It is made up of the histogram, signal

line, and MACD line. Using MACD, one may create buy or sell signals and spot possible trends.

3.4 Bollinger Bands

Bollinger bands are made up of an N-period simple moving average (SMA) as the middle band, an upper band located at K times the standard deviation of the N period above the middle band, and a lower band located at K times the standard deviation of the N period below the middle band. When determining volatility and possible overbought or oversold situations, Bollinger Bands are utilized.

The mathematical Formula used in calculating the Bollinger Bands are as follows

1. Middle Band (MA)

MA SMA(P, n), where SMA is the Simple Moving Average, P is the price data, and n is the number of periods.

2. Upper Band

MA+kxSD, where k is a constant (usually 2) and SD is the Standard Deviation of the price data.

3. Lower Band

Lower Band - MA-k x SD

Here, k determines the width of the bands, and SD is the standard deviation calculated over the same n periods

4. Comparison Results

The models are implemented using the Scala, both the model are trained using 75 % of yahoo finance dataset (2018 -2023) 25% of the dataset are used in testing the model's performance. The performance of DT and LSTM is evaluated using the metrics accuracy, precision, recall, and F1 score. The figure .3 below shows the effect of data points in the stock market prediction accuracy.

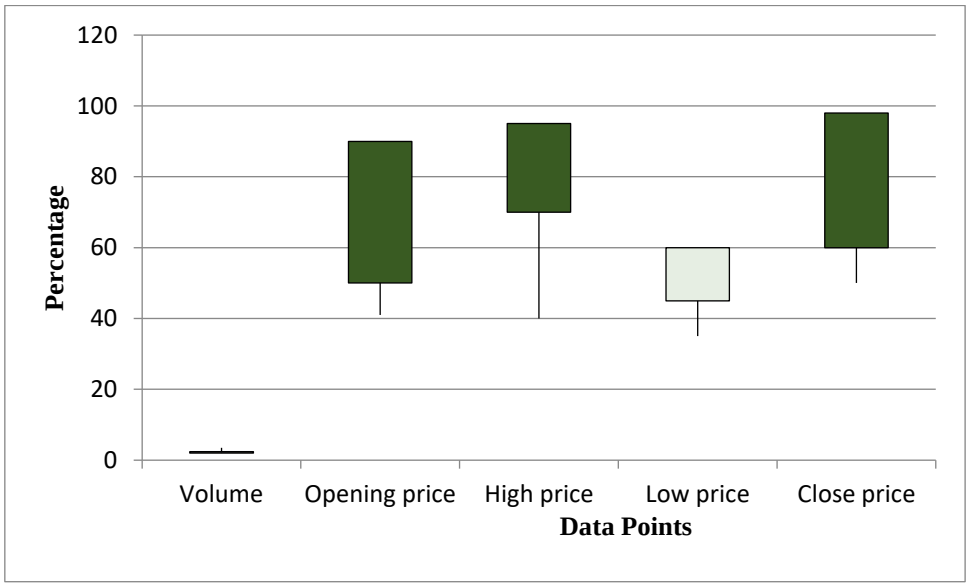

Figure 3. Effect of Data Points in Prediction Accuracy

The Table.2 illustrates the in-depth analysis of the output observed from the DT and the LSTM based on the training and the testing accuracy and loss

Table 2. Training and Testing Accuracy and Loss

Models	**Training accuracy**	**Training Loss**	**Testing Accuracy**	**Testing Loss**
Decision Tree	0.7445	0.3553	0.7123	0.3321
Long Short-Term Memory	0.8546	0.2134	0.8322	0.1845

The graphical representation below shows the performance of the two models DT and the LSTM. As the decision tree models does not have the capability of identifying the intricate pattern in time series data. The accuracy of the DT was observed to be slightly lesser compared to the LSTM. LSTM offered a better accuracy compared to the DT because of the large dataset used. However, LSTM was more complex and computationally intensive compared to the DT. So, from the study it was understood that difficulties in predicting the stock values due to its financial values being influenced by the various factors the combination of different models and the ensemble methods would be more optimal. In future the study would leap further with

the comparison of combination of models and ensemble methods to have better approach for developing an automated stock market predictor.

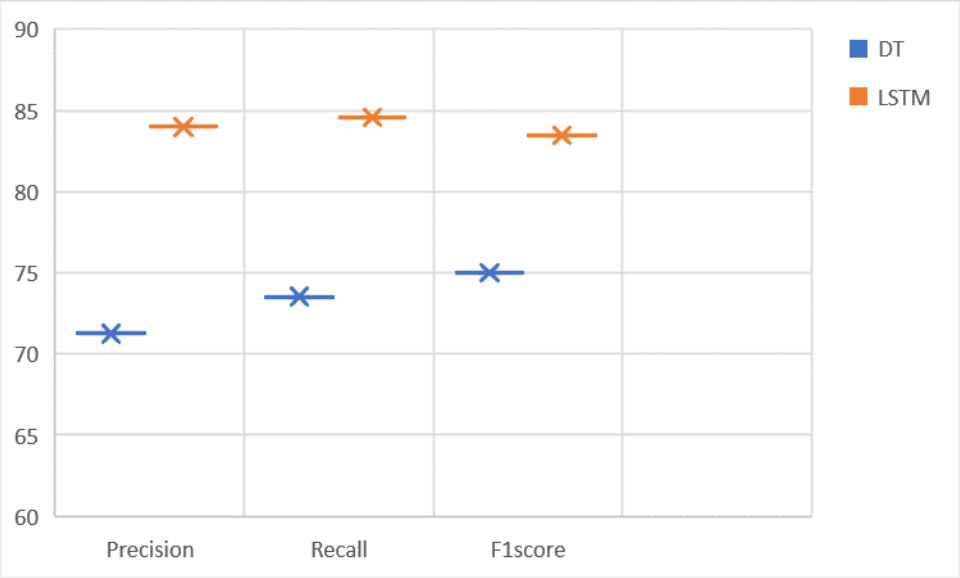

Figure 4. Performance of Decision Tree and LSTM

5. Conclusion

The proposed study compares the performance of the decision tree and LSTM in forecasting the stock market prices using the yahoo finance dataset. The models were trained and tested in Scala. The performance evaluation based on the accuracy, precision, recall and F1 score proved that the LSTM was better compared to DT but still had few drawbacks. To overcome the drawbacks and find the optimal model to develop an automated stock market predictor the study in future would move with the comparison of more hybrid models and ensemble methods

Reference

[1] Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. New York: Wiley; 2015.

[2] Huang, Chao, Li-li Huang, and Ting-ting Han. "Financial time series forecasting based on wavelet kernel support vector machine." In 2012 8th International Conference on Natural Computation, pp. 79-83. IEEE, 2012.

[3] Wu, Jui-Yu, and Chi-Jie Lu. "Computational intelligence approaches for stock price forecasting." In 2012 International Symposium on Computer, Consumer and Control, pp. 52-55. IEEE, 2012.

[4] Kao, Ling-Jing, Chih-Chou Chiu, Chi-Jie Lu, and Chih-Hsiang Chang. "A hybrid approach by integrating wavelet-based feature extraction with MARS and SVR for stock index forecasting." Decision Support Systems 54, no. 3 (2013): 1228-1244.

[5] Soni, Payal, Yogya Tewari, and Deepa Krishnan. "Machine Learning approaches in stock price prediction: A systematic review." In Journal of Physics: Conference Series, vol. 2161, no. 1, p. 012065. IOP Publishing, 2022.

[6] Mintarya, Latrisha N., Jeta NM Halim, Callista Angie, Said Achmad, and Aditya Kurniawan. "Machine learning approaches in stock market prediction: a systematic literature review." Procedia Computer Science 216 (2023): 96-102.

[7] Patel, Ramkrishna, Vikas Choudhary, Deepika Saxena, and Ashutosh Kumar Singh. "Review of stock prediction using machine learning techniques." In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 840-846. IEEE, 2021.

[8] Gandhmal, Dattatray P., and K. Kumar. “Systematic analysis and review of stock market prediction techniques.” Computer Science Review 34 (2019): 100190.

[9] Ray, Susmita. “A quick review of machine learning algorithms.” In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon), pp. 35-39. IEEE, 2019.

[10] Fischer, Thomas, and Christopher Krauss. "Deep learning with long short-term memory networks for financial market predictions." European journal of operational research 270, no. 2 (2018): 654-669.

[11] Leung, Carson Kai-Sang, Richard Kyle MacKinnon, and Yang Wang. "A machine learning approach for stock price prediction." In *Proceedings of the 18th International Database*

[12] https://www.kaggle.com/datasets/suruchiarora/yahoo-finance-dataset-2018-2023