International Journal of Multidisciplinary Research and Growth Evaluation.

Forecasting Nasdaq stock progressions using classification and deep learning techniques

Dr. Syed Shabbeer Ahmad [1*], Dr. Krishna Prasad K [2]
[1]Post-Doctoral Research Scholar (23SUPDR16), Institute of Engineering and Technology, Srinivas University, Mangalore, Karnataka, India Orchid Id: 0000-0003-2931-8218, Email Id: shabbeer.ahmad@mjcollege.ac.in
[2]Professor, HOD of Cyber Security and Cyber Forensics, Institute of Engineering and Technology, Srinivas University, Mangalore, India. Orchid Id: 0000-0003-1081-5820, Email Id: krishnaprasadkcci@srinivasuniversity.edu.in,

* Corresponding Author: **Dr. Syed Shabbeer Ahmad**

Article Info

ISSN (online): 2582-7138
Volume: 04
Issue: 05
September-October 2023
Received: 15-07-2023
Accepted: 12-08-2023
Page No: 40-49

Abstract
Stock Market prices have always been unpredictable resulting in a lot of risk for its investors. This proposal uses machine learning techniques (Decision Tree, Random Forest, Adaptive Boosting (Adaboost), eXtreme Support Vector Classifier (SVC), Logistic Regression and deep learning methods such as Long short-term memory (LSTM) to build modules that can be used to predict accurate stock prices reducing the chances of risk and increasing in gains. In this proposal the National Association of Securities Dealers Automatic Quotation System (NASDAQ) stock data is being used which has been extracted from Yahoo Finance to predict and analyze various Stock Progressions.

DOI: https://doi.org/10.54660/.IJMRGE.2023.4.5.40-49

Keywords: Forecasting, Stock, Prediction, Patterns, Treands

1. Introduction

Forecasting stock exchange prices is one of the complex problems that can be solved by machine learning and deep learning techniques. The stock data is always nonlinear and does not follow any particular pattern and is governed by multiple factors such as nature of markets, investor ideology, performance of various public companies, politics etc., which need complex mathematical calculations to understand and predict it. It is essential that the stock market information should be effectively and efficiently processed before it is fed to various machine learning models. Using this mechanism, the stock and index values can be predicted with greater accuracy. Stock market prediction system will help immensely to stop losses that might incur due to sudden falling up stock prices can benefit the traders and stockholders. The machine learning models have the capabilities to automatically identify and learn the patterns within the start data and hence are suitable for predicting stock trends.

In this proposal, the performance of various machine learning models like XGBoost, Adaboost, Linear regression etc and deep learning mechanisms like LSTM to forecast the stock trends. 10 different technical indicators for the past 10 years serves as the input to our system. Two different approaches for productions are also employed. They are non-discrete approach and discrete approach to understand the impact of preprocessing before feeding the data to the machine learning models. The discrete approach utilizes data such as open, close, high and low values of the stocks. The non-discrete approach utilizes pre-processing to convert non discrete data into discrete data and then feed it to the model for stock prediction. The performances thoroughly evaluated for three classification metrics and identify the best tuning parameter for each model. I strongly believe that this model can pave a way and help in building a real time system that can predict stock market trends in live environments.

2. Related Works

[1]" A Local and Global Event Sentiment Based Efficient Stock Exchange Forecasting Using Deep Learning"

The author Maqsood Haider, Mehmood has give a proposal of stock exchange using deep learning in the following manner. Stock exchange forecasting is an important aspect of business investment plans. The customers prefer to invest in stocks rather than traditional investments due to high profitability.

The high profit is often linked with high risk due to the nonlinear nature of data and complex economic rules. The stock markets are often volatile and change abruptly due to the economic conditions, political situation and major events for the country. Therefore, to investigate the effect of some major events more specifically global and local events for different top stock companies (country-wise) remains an open research area. In this study, we consider four countries- US, Hong Kong, Turkey, and Pakistan from developed, emerging and underdeveloped economies' list. We have explored the effect of different major events occurred during 2012–2016 on stock markets. We use the Twitter dataset to calculate the sentiment analysis for each of these events. The dataset consists of 11.42 million tweets that were used to determine the event sentiment. We have used linear regression, support vector regression and deep learning for stock exchange forecasting. The performance of the system is evaluated using the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The results show that performance improves by using the sentiment for these events.

[2]"Deep Learning Based Feature Engineering For Stock Price Movement Prediction"

The author M. A. Sheikh has give a proposal of stock exchange using deep learning in the following manner. Stock price modeling and prediction have been challenging objectives for researchers and speculators because of noisy and non-stationary characteristics of samples. With the growth in deep learning, the task of feature learning can be performed more effectively by purposely designed network. In this paper, we propose a novel end-to-end model named multi- filters neural network (MFNN) specifically for feature extraction on financial time series samples and price movement prediction task. Both convolutional and recurrent neurons are integrated to build the multi-filters structure, so that the information from different feature spaces and market views can be obtained. We apply our MFNN for extreme market prediction and signal-based trading simulation tasks on Chinese stock market index CSI 300. Experimental results show that our network outperforms traditional machine learning models, statistical models, and single-structure(convolutional, recurrent, and LSTM) networks in terms of the accuracy, profitability, and stability.

[3]"Stock Price Prediction using LSTM and SVR"

The author Gourav Bathla has given a proposal of stock prediction using LSTM and SVR.

Stock price movement is non-linear and complex. Several research works have been carried out to predict stock prices. Traditional approaches such as Linear Regression and Support Vector Regression were used but accuracy was not adequate. Researchers have tried to improve stock price prediction using ARIMA. Due to very high variations in stock prices, deep learning techniques are applied due to its proven accuracy in various analytics fields. Artificial Neural Network was deployed to predict stock prices but as stock prices are time-series based, recurrent neural network was applied to further improve prediction accuracy. In RNN, there is limitation of not able to store high dependencies and also vanishing gradient descent issue exists. Therefore, data scientists and analysts applied LSTM to predict stock price movement. In this paper, LSTM is compared with SVR using various stock index data such as S& P 500, NYSE, NSE, BSE, NASDAQ and Dow Jones industrial Average for experiment analysis. Experiment analysis proves that LSTM provides better accuracy as compared to SVR.

3. Problem Statement

The stock data is always nonlinear and does not follow any particular pattern and is governed by multiple factors such as nature of markets, investor ideology, performance of various public companies, politics etc., which need complex mathematical calculations to understand and predict it. It is essential that the stock market information should be effectively and efficiently processed before it is fed to various machine learning models. Using this mechanism, the stock and index values can be predicted with greater accuracy. Stock market prediction system will help immensely to stop losses that might incur due to sudden falling up stock prices can benefit the traders and stockholders.

The machine learning models have the capabilities to automatically identify and learn the patterns within the start data and hence are suitable for predicting stock trends. This system will help the stockholders and traders to make appropriate decisions and help in the process of stock trading. As the stock market information would contain multiple patterns, machine learning models are the best way to predict the stock trends and prices these models can learn the patterns by themselves. This would be of immense help as the traders can keep up with the fluctuating prices and improvise the trading decisions.

I aim to predict the trends with respect to this data and build a web application using which users can input the technical indicators of their choice and get predictions about the trend. Maintaining user history or working with live environments does not fall under the purview of this proposal.

4. Proposed Method

4.1 Objective

In this proposal, my objective is to come up with effective and efficient machine learning models that can predict the movement of prices in stock exchange accurately. This system will help the stockholders and traders to make appropriate decisions and help in the process of stock trading. As the stock market information would contain multiple patterns, machine learning models are the best way to predict the stock trends and prices these models can learn the patterns by themselves. This would be of immense help as the traders can keep up with the fluctuating prices and improvise the trading decisions.

4.2. Aim of the project

In this proposal, I aim to come up with effective and efficient machine learning models that can predict the movement of prices in stock exchange accurately. This system will help the stockholders and traders to make appropriate decisions and help in the process of stock

trading. As the stock market information would contain multiple patterns, machine learning models are the best way to predict the stock trends and prices these models can learn the patterns by themselves. This would be of immense help as the traders can keep up with the fluctuating prices and improvise the trading decisions.

4.3. Scope of the Project

I have initially done the data analysis for these IT companies using past one year data from Yfinance. The evaluation of algorithms and creation of model uses 10 years of stock data from Yfinance. I have truncated the last 60 days of data and used it for testing and prediction. The remaining data of 9 years and 10 months has been used for training the LSTM model. When the predictions are made on the test data which is the last 60 days data and compared with the actual stock values of the last 60 days, we observed that the LSTM model that I have developed has given high accuracy with very minimal deviation. The application is hosted as a web application where the admin of the system can analyse the stock data, compare multiple algorithms on the dataset and create the final model for predicting the stock values. The application is hosted as a web application for users to utilize the services of making future predictions for stock data.

4.4. Existing System

Stock price forecasting has always been a arduous issue for financial and statistical experts. There is a need to forecast the prices of stocks as one has to purchase stocks that might increase and sell stocks that might decrease in price in the near future so as to maximize the profits .minimize the losses. The traditional ways of forecasting stock prices are fundamental analysis which relies on in organization's annual growth rate, market position etc. The Other method is technical analysis mechanism and it relies on the stocks previous prices and values. It uses historical data and patterns to forecast the stock price. These methods do not accurately forecast the prices as there are many uncertain factors that directly influence the stock prices like political situations, company public profile, local and global events etc.

Drawbacks of Existing System The existing techniques cannot predict the future trends. Might incur heavy losses

4.5. Proposed System

In this proposal, emphasizes over evaluation process on prediction performance of five machine learning models (Decision Tree, Random Forest, Adaboost, SVC, Logistic Regression) and deep learning methods (LSTM) towards forecasting stock market trends. I have initially done the data analysis using past one year data of Microsoft corporation from Yfinance. However, the evaluation of algorithms and creation of model uses 10 years of stock data from Yfinance. I have truncated the last 60 days of data and used it for testing and prediction. The remaining data of 9 years and 10 months has been used for training the LSTM model. When the predictions are made on the test data which is the last 60 days data and compared with the actual stock values of the last 60 days, we observed that the LSTM model that we have developed has given high accuracy with very minimal deviation. The application is hosted as a web application where the admin of the system can analyse the stock data, compare multiple algorithms on the dataset and create the final model for predicting the stock values. The application is hosted as a web application for users to utilize the services of making future predictions for stock data. The application has been developed to predict the stock values for 4 top IT companies from such as Apple, Google, Amazon, Nvidia corporation, but the system is capable of forecasting the stock values for the next 60 days for any given ticker symbol.

Advantages:

Able to predict the stock market trends accurately.

Minimize loss and maximize profits

5. Result Analysis

5.1. Execution

The proposed modular implementation of the project consists of two modules:

Admin

User

Admin Module:

The admin of the system is responsible for the activities like:

Get stock data from yahoo finance

Data Analysis of the dataset

Splitting the dataset for training and testing

Training the model for multiple algorithms

Review the performance of the algorithms on the given dataset

Create the model using LSTM algorithm.

User Module:

The user of the system can utilize the machine learning services that are offered like:

Logging into the system

Enter stock details to predict future trends

Receive prediction for future trend

Home page:

This is the starting page of the application when the application is executed on Pycharm, the application is hosted on a web server and URL is generated to access the application once the user clicks on the URL the page represented in Figure 1 is opened in the browser.

Fig 1: Home Page

Admin Login:
This is the login page for the admin module. The admin need to login into the system with his credentials in order to perform operations like uploading the dataset, Training the dataset, Exploratory data Analysis of the dataset, Feeding the dataset to different Machine learning
Algorithms to find the Algorithm that can meet the best accuracy and Create a model that can be hosted on the Flask Application to be used by the users. Figure 2 represents the admin login page.

Fig 2: Admin Login

Upload Dataset:
On this page, the administrator of the system can upload datasets that are used for training the machine learning models. The admin has to select the file by clicking on the Choose file button and click on the upload button to upload the file to the server. Once the upload is complete, a success message would be displayed that the file is successfully uploaded. For this project we are using MSFT.csv (Microsoft) as a dataset. Figure 3 and Figure 4 shows the data being uploaded on the page.

Fig 3: Upload dataset

Fig 4: File Uploaded Successfully

Exploratory Data Analysis

Exploratory Data Analysis is performed on the dataset in order to clean the dataset for any missing data, identify patterns, identify the relationships of various parameters of the outputs with the help of graphs, statistics etc.

Price Analysis

Figure 5 shows the Price analysis graph of past one year Stock Data for Microsoft company from MSFT.csv File.

Fig 5: Price Analysis

Sales Volume Analysis

The Figure 6 shows the Sales Volume Analysis graph for past one year Stock Data of Microsoft company from MSFT.csv File.

Fig 6: Sales Volume Analysis

Daily Return Analysis

Figure 7 depicts the analysis for past one year Stock Data of Microsoft company from MSFT.csv File.

Fig 7: Daily Return Analysis

Moving Average Analysis

Figure 8 shows the Moving Average Analysis graph for Adjustment close of every day, Moving Average of 10 days, 20 days & 50 days of past one year Stock Data of Microsoft company from MSFT.csv File.

Fig 8: Moving Average Analysis

Logistic Regression

When the dataset is feed to Logistic regression algorithm, I observe that the test accuracy is 56.8627% represented in Figure 9.

Fig 9: Logistic Regression Results

Support Vector Machine

When the dataset is feed to Support Vector Machine algorithm, I observe that the test accuracy is 60.7843% represented in Figure 10.

Fig 10: Support Vector Machine Results

Decision Trees

When the dataset is feed to Decision Trees algorithm we observe that the test accuracy is 45.0980% represented in Figure 11.

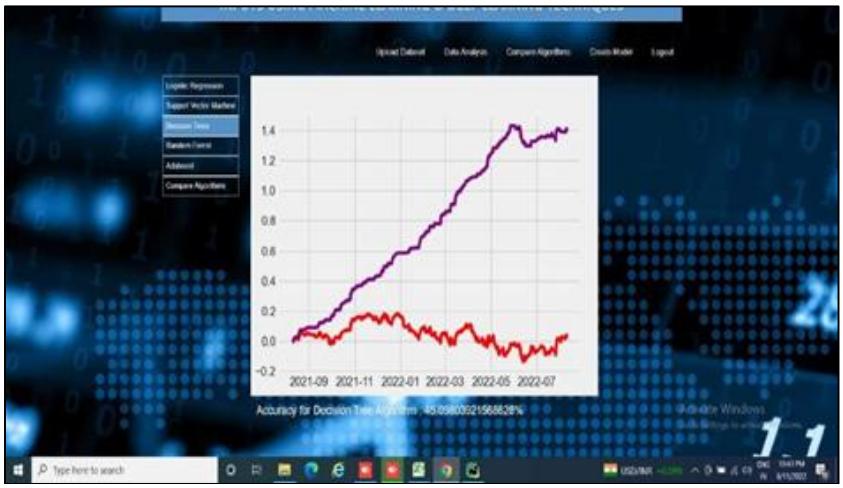

Fig 11: Decision Trees Results

Random Forest

When the dataset is feed to Random Forest algorithm weobserve that the test accuracy is 54.9019% represented in Figure 12.

Fig 12: Random Forest Results

Adaboost

When the dataset is feed to Adaboost algorithm we observe that the test accuracy is 50.9803%. represented in Figure 13.

Fig 13: Adaboost results

Compare All Algorithms

Figure 14 shows the comparison of various test accuracies of the Algorithms.

Fig 14: Comparing Analysis

Create Model

Figure15 shows the creation of Model for better optimized of system.

Fig 15: Create Model

5.2. Metrics of LSTM model (FinalClassifer)

Mse, mae, mape are Keras Regression Metrics
Mean Square Error (mse): 0.0013 Mean
Absolute Error(mae): 0.0211
Mean Absolute Percentage Error (mape): 10117.0127 Root mean square error: 7.942244122086709Metrics for Algorithms in CompAlg.py

Classification report contains the complete metric information of the evaluated algorithm. They are Precision, Recall, F1-Score, Support
Precision – What percent of your predictions were correct?
Precision is the ability of a classifier not to label an instance positive that is actually negative.For each class it is defined

as the ratio of truepositives to the sum of true and false positives.

TP – True Positives
FP-False Positive

Precision – Accuracy of positive predictions.
Precision = TP/(TP+FP)
Recall – What percent of the positive cases did you catch?
Recall is the ability of a classifier to find all positive instances. For each class it is defined as the ratio of true positives to the sum of true positives and false negatives.
FN – False Negatives
Recall : Fraction of positives the were correctly identified
Recall = TP/(TP+FN)
F1 score – What percent of positive predictions were correct?
The F1 score is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst is 0.0. Generally speaking, F1 scores are lower than accuracy measures as they embed precision and recall into their computation. As a rule of thumb, the weighted average of F1 should be used to compare classifier models, not global accuracy.
F1 Score = 2*(Recall * Precision) / (Recall + Precision)
Support is the number of actual occurrences of the class in the specified dataset. Imbalanced support in the training data may indicate structural weaknesses in the reported scores of the classifier and could indicate the need for stratified sampling or rebalancing.

SVM Accuracy: 25.49019607843137
SVM Classification Report:
Precisionrecallf1-scoresupport
00.30 0.09 0.14 34
10.24 0.59 0.34 17
Accuracy0.2551
Macro Average 0.27 0.34 0.2451
Weighted avg 0.28 0.25 0.21 51

Logistic Regression Accuracy: 58.82352941176471
Logistic Regression Classification Report:
Precisionrecallf1-scoresupport

00.71 0.65 0.68 34
10.40 0.47 0.43 17
Accuracy0.5951
Macro avg0.55 0.56 0.55 51
Weighted Avg0.610.590.6051
Random Forest Accuracy: 41.17647058823529
Random Forest Classification Report:
Precisionrecall f1-score support
00.590.380.46 34
10.280.470.35 17
Accuracy0.4151
Macro avg0.43 0.43 0.41 51
Weighted Avg0.490.410.4351
Decision Tree Accuracy: 41.17647058823529
Decision Tree Accuracy Report
Precisionrecall f1-score support
00.58 0.41 0.48 34
10.26 0.41 0.32 17
Accuracy0.4151
Macro avg0.42 0.41 0.40 51
Weighted Avg0.480.410.4351
Adaboost Accuracy: 49.01960784313725
Adaboost Tree Accuracy Report
Precisionrecall f1-score support
00.64 0.53 0.58 34
10.30 0.41 0.35 17
Accuracy0.4951
macroavg0.47 0.47 0.47 51
Weighted Avg0.530.490.5051

The results of the predicted stock prices are shown in Figure 17.

Fig 16: Prediction Stock Prices

6. Conclusion

In this proposal we successfully predicted the stock market trends using deep learning and machine learning techniques.10 years of stock data of top IT companies listed on NASDAQ from Yfinance such as Microsoft. Comparison of five machine learning models (Decision Tree, Random Forest, Adaptive Boosting (Adaboost), eXtreme Support Vector Classifier (SVC), Logistic Regression and deep learning methods such as Long short-term memory (LSTM). I have initially done the data analysis for these IT companies using past one year data from Yfinance. The evaluation of algorithms and creation of model uses 10 years of stock data from Yfinance. We have truncated the last 60 days of data and used it for testing and prediction.

The remaining data of 9 years and 10 months has been used for training the LSTM model. When the predictions are made on the test data which is the last 60 days data and compared with the actual stock values of the last 60 days, we observed that the LSTM model that we have developed has given high accuracy with very minimal deviation The application is hosted as a web application for users to utilize the services of making future predictions for stock data. The application has been developed to predict the stock values for 4 IT companies but the system is capable of forecasting the stock values for the next 60 days for any given ticker symbol and achieved about 95% of accuracy.

In future, the UI of the system can further be enhanced to make stock predictions for any ticker and also predict stock values of stocks from other stock exchanges other than NASDAQ.

7. References

1. Maqsood H, Mehmood I, Maqsood M, Yasir M, Afzal S, Aadil F, *et al*. A local and global event sentiment based efficient stock exchange forecasting using deep learning. International Journal of Information Management. 2020;50:432-51.
2. Long W, Lu Z, Cui L. Deep learning-based feature engineering for stock price movement prediction. Knowledge-Based Systems. 2019;164:163-73.
3. Bathla G. Stock price prediction using LSTM and SVR. In: Proceedings of the 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC); c2020.
4. Saud AS, Shakya S. Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE. Procedia Computer Science. 2020;167:200-208.
5. Nabipour M, Nayyeri P, Jabani H, Mosavi A. Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data: A comparative analysis. IEEE Access. 2019;8:150199-212.
6. Eapen J, Bein D, Verma A. Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction. In: Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC); c2019.
7. Duarte JB, Duarte LH, Talero Sarmiento KJ, Sierra Juárez K. Evaluation of the effect of investor psychology on an artificial stock market through its degree of efficiency. Contaduría y Administración. 2017;62(4):1361-1376.
8. Lu N. A Machine Learning Approach to Automated Trading. Boston, MA, USA: Boston College Computer Science Senior; c2016.
9. Jiang W. Applications of deep learning in stock market prediction: Recent progress. Expert Systems with Applications. 2021;184:115264.
10. Pang X, Zhou Y, Wang P, *et al*. An innovative neural network approach for stock market prediction. Journal of Supercomputing. 2020;76:2098-2118.
11. Thakkar A, Chaudhari K. Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions. Information Fusion. 2021;65:1-22.
12. Thakkar A, Chaudhari K. Cross-Reference to Exchange-based Stock Trend Prediction using Long Short-Term Memory. Procedia Computer Science. . 2020;167:616-625.
13. Zhang C, Sjarif NNA, Ibrahim RB. Decision fusion for stock market prediction: A systematic review. IEEE Access. 2022;10:81364-379.
14. Li AW, Bastos GS. Stock market forecasting using deep learning and technical analysis: A systematic review. IEEE Access. 2020;8:185232-185242.