Prediction of Exact Price of Stock and Direction of Stock Market using Statistical and LSTM Model

Atul Mishra
School of Engineering & Technology
BML Munjal University
Gurgaon, India
atul.mishra@bmu.edu.in

Ranbir Singh
School of Engineering & Technology
BML Munjal University
Gurgaon, India
ranbir.singh@bmu.edu.in

Anubhav Agrawal
School of Engineering & Technology
BML Munjal University
Gurgaon, India
anubhav.agrawal@bmu.edu.in

Pradeep Kumar Arya*
School of Computer Science Engineering and Technology,
Bennett University
Grater Noida, UP, 201310, India
pradeeparya25@gmail.com

Aditi Sharma*
Department of CSE, Symbiosis Institute of Technology,
Symbiosis International (Deemed University),
Pune,Maharastra,India
aditi.sharma@ieee.org

Abstract— **Global stock markets reflect significant changes in national economies, and as such, they attract a great deal of investor attention. The movements in the market are based on previous market changes. Predicting the stock market's direction is a crucial factor that affects financial investors' decisions to buy, hold, or sell shares. Therefore, investors are always concerned about analyzing and projecting long-term stock market trends to maximize their earnings. This paper examines various models used to replicate the stock market's behaviour, and a comparison has been made. The models used include Statistical Analysis, Regression Analysis, SVM, and LSTM. The study also analyzes the impact of Twitter sentiment on the stock market. The proposed approach has been evaluated using RELIANCE and RCOM share price data. The experimental results indicate that the proposed system using LSTM performs better than the state-of-the-art method.**

Keywords*— *Forecasting, regression, prediction, LSTM, financial

I. INTRODUCTION

Research on the impact of social media and its influence on the stock market is a popular topic, particularly when it comes to data and its application to social media. Stock market forecasting[1] has always been a subject of interest for academics, who try to predict whether prices will rise or fall. With the advancement of machine learning and neural network techniques, there has been a growing interest in this field. The focus now is on anticipating the direction of stock prices accurately, so investors can make informed decisions.

In recent years, numerous attempts have been made to predict stock market behavior. However, accurately representing the market's dynamics is challenging, given the stochastic process behavior of the exchange with time. Due to the dynamism of the markets involved, it's often challenging for a technique to consistently and correctly predict market variations. Nonetheless, these models have achieved satisfactory levels of success. Forecasting involves predicting a set of output variables based on input variables. In stock market forecasts, input variables may include a record of stock prices, investors' sentiments, industry performance, and others linked to the company's portfolio. The output variable will mainly comprise the stock market behavior in the future (i.e., whether prices will increase or decrease) [2].

The goal is to demonstrate a machine learning model that can predict future stock market behaviour (i.e., whether prices will rise or fall) with the least amount of error possible and present a model that will assist investors more effectively, according to the researchers. Because of the rapid expansion of the stock market trading business, the capacity to forecast a sequence of data points, such as stock prices, has become increasingly popular among financial analysts. Researchers are constantly being driven to look for new approaches and ideas that they may utilise to make their forecasts more accurate. Researchers in finance and investors and anyone participating in the securities market or share trading would benefit from developing any system or methodology that might be used to forecast stock price movements.

Stock markets across the globe imitate the current vacillations of affairs within the market economies of respective countries and, hence, are chargeable with the attraction of an excessive number of investors. The movements in available markets rely upon past fluctuations within the securities market. Prediction of the direction of stock markets' campaigns is the most crucial issue that influences financial investors' decisions to shop for, sell, or hold stakes in stock [3]. Because the stock market helps investors garner high profits, it also involves high risks. Hence, investors are always concerned about analysing and forecasting the longer-term trends of the stock markets to garner maximum profits. A summary of current studies investigating the influence of user sentiment on predicting the behaviour of the stock market is presented in the next section of the paper.

This article contains 5 sections. The introduction in discussed in Section 1 followed by review of recent trends in stock price prediction and stock movement in Section 2. In Section 3, various methods are discussed. Section 4 presents a comparative analysis of the performance measures associated with the strategies outlined in the preceding sections. Finally, Section 5 contains the conclusion and scope.

979-8-3503-7617-3/24/$31.00 ©2024 IEEE
DOI 10.1109/AIARS63200.2024.00184

II. LITERATURE REVIEW

This research explores a generic stock price prediction system that takes Twitter tweets as inputs and generates projected stock price movements as outputs. Recent trends highlight and explain forecasts made by directly implementing the consequences contained within the original trading-related acts. Neural networks are used to predict financial markets in terms of analysis and research, and they are becoming increasingly popular due to their versatility. For futures traders and stockbrokers to make trading judgments, they have relied on a range of various types of intelligent systems during their careers. In recent years, artificial neural networks (ANNs) have been used to model and predict the value of stocks [4, 5, 6, 7]. As a result of the vast amount of noise and advanced complexity of stock price data, these models have their limitations. On top of that, both the input elements and the quantity of knowledge itself may conflict with one another.

Brown et al. [5] explained the variation in the prediction of stock prices, and the movement in prices behaves as a stochastic process and varies with time. Generally, there are two analytical models used to predict the behaviour of the securities market. These are fundamental and technical analyses. Fundamental analysis includes industry analysis, economic analysis, and company analysis. Technical analysis could be a method that predicts future prices supported by past market data.

To support the idea of return on equity, Gupta, UL [6] indicates in his book that the stated rate of return on the new issues will substantially impact the present buyers who are authorised to have the factor of "fixed pricing." According to the findings of a recent study, the investment performance of current difficulties with equity shares, particularly those of new companies, merits more investigation. This study provides reliable estimates of the rates of return on equities, as well as an examination of the variability of such rates of return over time.

Kohli et. al [7] discovered that the prices of economic assets are non-linear, dynamic, and chaotic; thus, financial statistics are difficult to predict. Among the newest techniques, machine learning models are some of the most researched, given their capabilities for recognising complex patterns in various applications.

Predicting stock price movement every day could be a difficult task due to the variations within the financial market. Factors such as the sensitivity of the BSE and national indices must not be used to analyse the correlation on the various stock returns provided monthly from January 1982 to December 1993 [8].

Ryaly et al. [9] analysed the behaviour of the daily and weekly returns of 5 Indian securities market indices for the stochastic process from April 1996 to June 2001. In his analysis, Sherstinsky [10] identified new opportunities to improve the LSTM system and incorporated these extensions into the Vanilla LSTM network, resulting in the most general LSTM variant to date. The target reader has already been exposed to RNNs and LSTM networks. However, in recent research, evidence favouring the LSTM network and its parent, RNN, is undeniably reported while the training formulas are omitted altogether.

A summary of current studies investigating the influence of user sentiment on predicting the behaviour of the stock market is presented in the next section of the paper. Pathak & Nisha [11] represented a hybrid approach where they combined machine learning modules and sentiment analysis modules and trained the model on Indian news data for each stock. Their research shows that the Ridge Regressor gave the maximum accuracy of 85.4% among all the algorithms applied.

A recent study by Xiao et al. examined distinct features for different time frames due to complex factors. It used a novel combined approach that combined singular spectrum analysis and supports vector machines. Research demonstrates the best prediction performance based on four criteria, indicating that the process significantly impacts the prediction of stock prices.

III. METHODOLOGY

Our approach demonstrates a machine learning model where the most critical issues affecting the stock prices have been taken into consideration very carefully. We have taken 10-year historical data for the stock prices from trusted sources like Google Finance and portals of major Indian stock markets (NSE & BSE) [14].

The proposed methodology is as follows:

• *Selection of input variables:* The most critical factors affecting stock prices have been done. The factors included in the prediction are previous 5-day stock prices, Volatility of the stock and Volume of the store.

• Data Pre-processing: The dataset taken from the NSE website had been pre-processed to clean the dataset to get some preliminary insights.

• *Designing the architecture of the machine learning model:* The model's architecture was decided from the insights gathered from the pre-processing of data. The proposed methodology for Machine Learning Models is depicted in Fig. 1.

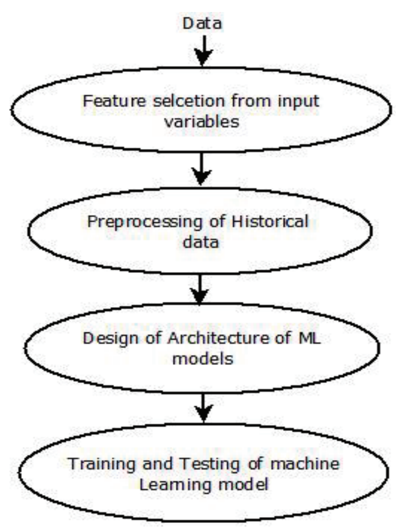


Figure 1: Proposed methodology for machine learning models

A. Market data

The first source of information comprises information on the stock's price returns, which is updated daily. We extract the time series of daily returns for each sock, which Rd denotes:

$$Rd = \frac{P(d) - P(d-1)}{P(d-1)}$$

Where Pd is the stock's closing price at day d, we employ raw returns rather than the more common log returns to maintain consistency with the original "event study." [15]. This data is publicly available and can be downloaded from various sources on the Internet [14]

B. Architecture

1. *Regression Model [13]:* We applied the same approach here also, where the last five days of historical data have been used to build the Linear Regression Model, and the coefficients have been calculated, which have been used for the prediction of future prices. The error calculated has been the Mean Absolute Error. The performance metrics are demonstrated accurately because the model is used to replicate the stock market's direction for the upcoming days.

2. *Support Vector Machines [12]:* SVM is best used for the time-series prediction and is considered as one of the optimization method. The application of SVM has been monitored both for regression and classification conditions. The SVM plots data as points in an n-dimensional space. These dimensions represent attributes that are plotted on coordinates. The SVM algorithm draws a boundary over the hyper-plane dataset, dividing data into two classes.

We applied support vector machines and followed the same approach for predicting future prices. The cost function used here changes accordingly, and the performance is reported using metrics such as accuracy, precision, and recall.

3. *Long-Short Term Memory:* It is a type of recurrent neural network, also known as an RNN, used to anticipate sequences of events. Noise resistance, trainable parameters, and the capacity to retain information for an arbitrary amount of time are the three most essential prerequisites for executing LSTM. In this model, the input for the input layer is made up of the last five days of historical data and the volume of stock on hand. To anticipate the outcome, the model makes use of 100 hidden layers. This output is transmitted to the dense layer, and the result acquired is then used to anticipate the stock market's direction. The evaluation of performance metrics takes place.

IV. Performance

The data was collected and developed to be converted into a shape that could be used as inputs in the model. The models built have used two different types of prediction variables: The exact price of a stock and the direction of the stock market, as depicted in Tables I and II.

TABLE I. PREDICTION ACCURACY AND MEAN ABSOLUTE ERROR

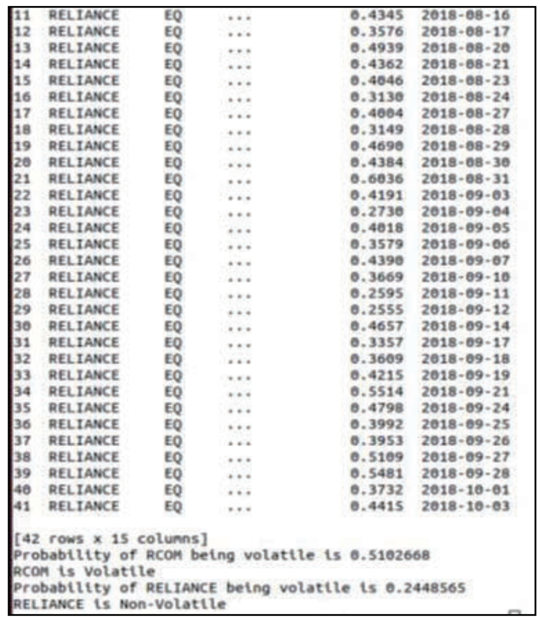
Prediction of Exact Price	Accuracy	Mean Absolute Error
Statistical Analysis	0.51	1% - 2%
Regression Analysis	0.51	1% - 3%

TABLE II. PERFORMANCE OF DIRECTION OF STOCK

The Direction of the Stock Market	Accuracy
Support Vector Machine	0.52
Long-Short Term Memory	0.62

Several models have performed differently in terms of stock market volatility for the stocks of RCOM and RELIANCE, as shown in Figs. 2 and 3. The comparison of stock prices is displayed in Fig. 3 using a statistical method to demonstrate the volatility character of the market.

Figure 2: Graph for stocks showing volatile nature



```
11  RELIANCE    EQ    ...    0.4345  2018-08-16
12  RELIANCE    EQ    ...    0.3576  2018-08-17
13  RELIANCE    EQ    ...    0.4939  2018-08-20
14  RELIANCE    EQ    ...    0.4362  2018-08-21
15  RELIANCE    EQ    ...    0.4046  2018-08-23
16  RELIANCE    EQ    ...    0.3130  2018-08-24
17  RELIANCE    EQ    ...    0.4004  2018-08-27
18  RELIANCE    EQ    ...    0.3149  2018-08-28
19  RELIANCE    EQ    ...    0.4690  2018-08-29
20  RELIANCE    EQ    ...    0.4384  2018-08-30
21  RELIANCE    EQ    ...    0.6036  2018-08-31
22  RELIANCE    EQ    ...    0.4191  2018-09-03
23  RELIANCE    EQ    ...    0.2730  2018-09-04
24  RELIANCE    EQ    ...    0.4018  2018-09-05
25  RELIANCE    EQ    ...    0.3579  2018-09-06
26  RELIANCE    EQ    ...    0.4390  2018-09-07
27  RELIANCE    EQ    ...    0.3669  2018-09-10
28  RELIANCE    EQ    ...    0.2595  2018-09-11
29  RELIANCE    EQ    ...    0.2555  2018-09-12
30  RELIANCE    EQ    ...    0.4657  2018-09-14
31  RELIANCE    EQ    ...    0.3357  2018-09-17
32  RELIANCE    EQ    ...    0.3609  2018-09-18
33  RELIANCE    EQ    ...    0.4215  2018-09-19
34  RELIANCE    EQ    ...    0.5514  2018-09-21
35  RELIANCE    EQ    ...    0.4798  2018-09-24
36  RELIANCE    EQ    ...    0.3992  2018-09-25
37  RELIANCE    EQ    ...    0.3953  2018-09-26
38  RELIANCE    EQ    ...    0.5109  2018-09-27
39  RELIANCE    EQ    ...    0.5481  2018-09-28
40  RELIANCE    EQ    ...    0.3732  2018-10-01
41  RELIANCE    EQ    ...    0.4415  2018-10-03

[42 rows x 15 columns]
Probability of RCOM being volatile is 0.5102668
RCOM is Volatile
Probability of RELIANCE being volatile is 0.2448565
RELIANCE is Non-Volatile
```

Figure 3: Probability for stock showing a volatile nature

The statistical and regression models are used to analyse the price prediction accuracy. The comparison between the actual and forecasted stock prices is depicted in Fig. 4 and Fig. 5. The graph demonstrates that the forecast is pretty accurate in both instances.

Figure 4: Prediction of stock prices for Reliance Communication (RCOM)

Figure 5: Prediction of stock prices for Reliance Industries (RELIANCE)

Figures 6 and 7 depict a comparison between accuracy and the requisite number of days for forecasting RCOM and RELIANCE shares. The analysis reveals that utilizing five days of historical data optimally informs model creation.

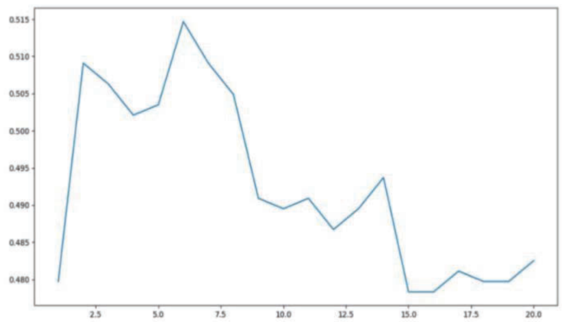


Figure 6: Accuracy of prediction with number of days as input for RCOM

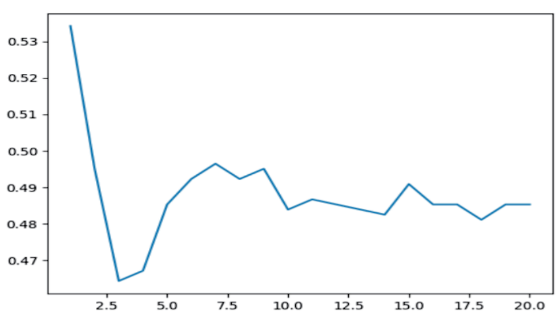


Figure 7: Accuracy of prediction with number of days as input for RELIANCE

Fig. 8 is the graph we obtained for predicted stock prices using pre-historic data on the LSTM model.

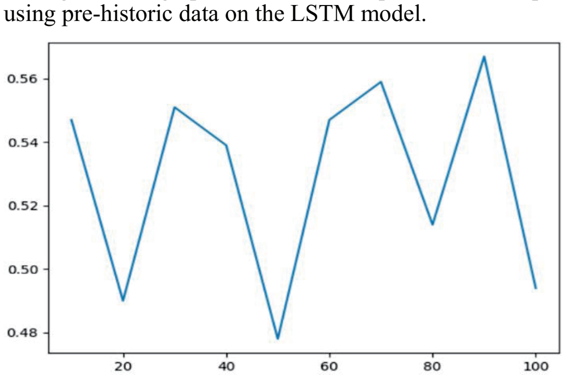


Figure 8: Accuracy of prediction with number of input layers in LSTM

Figure 8 illustrates the forecasted stock prices derived from an LSTM model trained on historical data. Our analysis suggests that employing 90 hidden layers optimally balances prediction accuracy. We assess model performance using the R2 score, indicating the alignment of data with the regression line. Notably, the similarity between test and train set values suggests absence of overfitting or underfitting. To validate our findings, 0.2% of the training data was reserved.

Figure 9 depicts the loss incurred during training and testing phases. The R2 scores for the training and testing datasets stand at 0.98 and 0.92, respectively.2

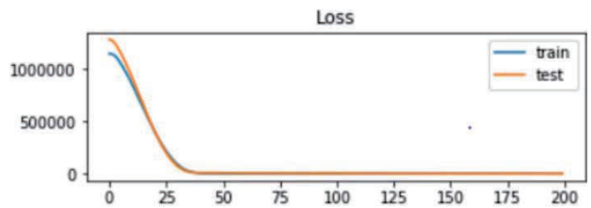


Figure 9: Comparison of Loss in test and train data

V. Conclusion

In conclusion, the pursuit of forecasting techniques for stock prices has become crucial in response to the rapid growth of the stock market. Our study employed time-series data to develop a model yielding significant results. This approach holds promise for application in other sectors, such as banking. Future improvements include incorporating sentiment analysis, and long-term prediction capabilities, extending the model to other data types, and integrating real-time data sources for enhanced accuracy. These advancements aim to empower decision-makers with actionable insights in navigating the complexities of the financial landscape.

References

[1] Pawar, Kriti, Raj Srujan Jalem, and Vivek Tiwari. "Stock Market Price Prediction Using LSTM RNN." Emerging Trends in Expert Applications and Security. Springer, Singapore, 2019. 493-503.

[2] Metawa, Noura, et al. "Impact of behavioural factors on investors' financial decisions: case of the Egyptian stock market." International Journal of Islamic and Middle Eastern Finance and Management (2019).

[3] Basak, Suryoday, et al. "Predicting the direction of stock market prices using tree-based classifiers." The North American Journal of Economics and Finance 47 (2019): 552-567.

[4] D'Ecclesia, Rita Laura, and Daniele Clementi. "Volatility in the stock market: ANN versus parametric models." Annals of Operations Research (2019): 1-27.

[5] Brown, S. J., Goetzmann, W. N., & Kumar, A. (1998). The Dow theory: William Peter Hamilton's track record reconsidered. The Journal of finance, 53(4), 1311-1333.

[6] Book by Gupta, U. L. (1972). Working of Stock Exchanges in India.

[7] Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. Expert Systems with Applications, 124, 226-251.

[8] Chopra, N. (2019). Sensitivity Analysis Using Garch Model: Evidence from Indian Stock Market. Journal of Commerce & Accounting Research, 8(2).

[9] Ryaly, V. R., Raju, G. S., & Urlankula, B. (2017). Testing the Weak-Form Market Efficiency in the Indian Stock Market: Evidence from the Bombay Stock Exchange Index (BSE) Sensex. Indian Journal of Finance, 11, 26-40.

[10] Sherstinsky, A. (2018). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. arXiv preprint arXiv:1808.03314.

[11] Pathak, A., & Shetty, N. P. (2019). Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis. In Computational Intelligence in Data Mining (pp. 595-603). Springer, Singapore.

[12] Usmani, M., Adil, S. H., Raza, K., & Ali, S. S. A. (2016, August). Stock market prediction using machine learning techniques. In 2016 3rd International Conference on Computer and Information Sciences (ICCOINS) (pp. 322-327). IEEE.

[13] Ihlayyel, Hani AK, Nurfadhlina Mohd Sharef, and Mohd Zakree Ahmed Nazri. "An enhanced feature representation based on linear regression model for stock market prediction." Intelligent Data Analysis 22.1 (2018): 45-76.

[14] NSE https://www.nseindia.com

[15] Campbell JY, Lo AWC, MacKinlay . The econometrics of financial markets. Princeton University Press; 1997.