Machine-learning-based prediction of stock market prices

Mr. Uday Joshi, Prof. Rashmi Tundalwar , Dr. Arati Dandavate

1ME Student Computer Science and Engineering Dhole Patil College of Engineering, Pune,
2 Guide Computer Science and Engineering Dhole Patil College of Engineering, Pune,
3Head Of Department Computer Science and Engineering Dhole Patil College of Engineering, Pune,

Abstract▬ Investment corporations, hedge funds, and even individual investors must study financial models to understand market behavior and build profitable investments and trades. Investors usually produce educated guesses by analyzing information about old stock prices, the company's performance behavior, etc. The initial phase of revealing theories in the guesswork indicates that stock unit prices are entirely random and unpredictable. In the betterment of the guesswork, quantitative analysts get deployed to make prophetic models. The paper focuses on using machine learning techniques to develop better models for enabling appropriate recommendations for financial investments.

.

Index Terms- Stock Price Prediction, Machine Learning, Random Forest Regression.

I. INTRODUCTION

The world's stock markets comprehend enormous wealth. As with the extended market, investors hunted for ways to amass data regarding the companies listed in the market.

In the past, investors relied upon their expertise to spot market patterns, but this is not possible nowadays. Easily applied math analysis of financial information provides some insights. However, in recent years, investment firms have used numerous artificial intelligence (AI) systems to look for patterns in vast amounts of real-time equity and financial information. These systems support investment decision-making, and they have currently been used for a

The sufficiently long amount that their features and performance will be reviewed and analyzed to

Identify those systems and improve prophetic performance compared with alternative techniques.

When the prediction goes correct, the vendor and stock broker make enormous profits. Frequently when the prediction goes in unexpected ways it is expected by analyzing the history of several securities markets. Machine learning is economical, thanks to representing such processes. It predicts a market price value near the physical weight with increasing accuracy—the introduction of machine letypes of research attributable to its economic and correct values measurements.

Dataset is the important part of machine learning used in education. The information set ought to be as concrete as potential, resulting from which amendment within the data will uphold massive changes within the outcome. This project uses supervised machine learning on a dataset obtained from Yahoo Finance. This dataset has five variables: open, close, low, high, and volume. With nearly direct names, airy, compact, soft, and increased area units indicate different bid costs at other times. Throughout the fundamental measure, shares are passed from one owner to another. The test information is then used to develop a model. A regression model and an LSTM model are used to test this conjecture, one by one. During working hours, regression minimizes errors, and LSTM contributes to the cognitive process of information and result. Last but not least, graphs for the fluctuation of cost with dates (for the regression-based model) and between actual and expected prices (for the LSTM-based model) are planned.

Stock Market Prediction aims to predict the longer-term price of a corporation's money stocks. Market prediction technologies use machine learning to make predictions based on current exchange indices and coaching on previous values. By employing different models, machine

learning creates more accurate and detailed forecasts. Our primary focus is on utilizing regression and LSTM machine learning techniques and developing a deep understanding of stock values. Several factors are considered, including the area of the unit, the low, the high, and the volume of stock values..

This Paper introduces several techniques for calculating the prices like the R factor, Quantitative Analysis.

R factor - The chance/praise ratio, often called the "R/R ratio," compares the capability income of a change to its capability loss. It is a calculation that uses the distinction between the access factor of a difference and the stop-loss to decide chance and the distinction between the income goal and the access factor to locate praise.

Quantitative Analysis - Quantitative evaluation (QA) in finance is a technique that emphasizes mathematical and statistical evaluation to assist decides the price of a monetary asset, along with an inventory or option. Quantitative buying and selling analysts (additionally recognized as "quanta") use several data—including historical funding and inventory marketplace data—to increase buying and selling algorithms and pc models.

II. Related Work

Stock price prediction can be predicted using AI and machine learning models in machine learning fields. It uses the SVM model for stock price prediction. Support vector machine which works on classification algorithms. It is used to get a new text as an output. Applying Multiple Linear Regression with Interactions to predict the trend in stock [1]

Using data from stock markets around the globe, Beginner's checks whether the markets are efficient and whether there are any anomalies. Whenever a market anomaly is found, scholars first confirm the anomaly and then search for existing models to explain the anomaly. Suppose scholars are unable to estimate, evaluate, and forecast any model to explain the anomaly. In that case, scholars will use quantitative analysis, modeling, or even a new theory of information to explain the anomaly that led to Behavioral Finance. In the event of an unexplained anomaly, one may be able to exploit the monster in order to profit. Investors can get valuable investment advice this way, on the one hand [2]The real Gross Domestic Product reflects the relationship between the stock market and the economic activity of the five European countries: Germany, France, Italy, the Netherlands, and the UK. This analysis includes variables such as stock market returns, actual economic activity, and interest rates in addition to the variables commonly used in such analyses. In the empirical VAR model, the authors have included the composite leading indicator [3].

The weak-form potency and stochastic process behavior of the CIVETS stock markets throughout the amount 2002–2012. We tend to apply unit root tests, serial autocorrelation, and variance quantitative relation tests. Our unit root results imply that CIVETS follow a stochastic process [4].

To predict the stock value of NSE and securities markets, two leading stock markets worldwide, the authors use four-deciliter architectures. We tend to train four networks, MLP, RNN, LSTM, and CNN, with the stock value of TATA MOTORS from NSE. From the NSE stock exchange, the models were used to predict the stock values of MARUTI, HCL, and AXIS BANK, and from the securities market, BANK OF AMERICA (BAC) and CHESAPEAKE ENERGY (CHK). Based on the results obtained, it is clear that the models can describe the patterns found in each stock market [5]. The importance of predicting the securities exchange price is well known among financial specialists since they need to know what kind of return they will receive for their investments. Generally, specialized experts and intermediaries use chronicled prices, volumes, value designs, and basic patterns to predict stock costs. Stock value expectations today are even more baffling than before since the organization's money-related status, as well as the socio-practical state of the nation, political environment, and cataclysmic events, influence stock costs. [6].

III. Proposed System

This paper introduced LSTM (Long Short Term Memory) model in stream-lit, which will predict the values based on the old dataset. The Prediction values are High, Low, Open, and Close. It is a reliable application for students and beginners who want to trade. They can quickly identify the trends in the market, whether the market is going upward or downward, or else it will remain sideways.

The model generates the confusion matrix for the classification report. This paper introduced the two regression and classification methods for stock market prediction. In the regression method, the closing price of company stock is predicted. The classification method will predict company stocks' closing price that will increase or decrease in upcoming days.
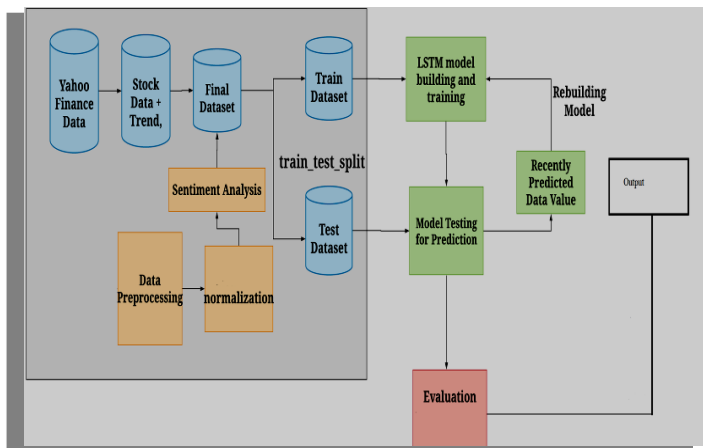
Proposed Architecture:

Fig.1. Proposed Architecture

Figure1 shows the proposed system design—this paper takes datasets from yahoo finance data. The first step is to train the data, and in the second step, data was tested, and with the LSTM model, forecast the values to get the prediction value.

The objective of the System

1. Explore Stock Prices.
2. Implement Basic Model Using Linear Regression
3. Implement LSTM Model with confusion matrix and classification report.

Method of Implementation

1. R Factor

   There are two types of equity market risks: systematic and non-systematic. Rising oil prices, currency movements, and changing government policies are familiar sources of frequent hazards. Unsystematic risks, however, are caused by factors unique to a company or industry. In addition, management and labor relations, increased competition, the entry of competing players, and customers' preference for a company's products all contribute to unsystematic risk.

2. **Stock Analysis Candle Stick Chart**

Candlestick charts show price movements of securities, derivatives, and currencies. As with a graph, each candle represents all four significant pieces of information for that day: open and close in the thick body; high and low in the wick. Two ways can be used to visualize buying and selling pressure using candlesticks.

3. LSTM Model

Long Short-Term Memory fashions are extraordinarily effective time-collection fashions. They can expect an arbitrary wide variety of steps into destiny. An LSTM module (or molecular) has five essential additives which permit it to version each long-time period and quick-time period data.

Cell nation (ct) - This represents the inner reminiscence of the molecular, which shops each quick period of reminiscence and long-time period recollections

Hidden nation (ht) - This is output nation records calculated w.r.t. modern enter, preceding remote country, and current molecular enter that you use to expect the destiny inventory marketplace prices. Additionally, the hidden nation can determine to handiest retrieve the short or long-time period or each variety of reminiscence saved withinside the molecular country to make the following prediction.

Input gate (it) - Decides how many records from current enter flow to the molecular nation.

Forget gate (ft) - Decides how many records from the modern enter and the preceding molecular nation flows into the contemporary molecular country.

Output gate (ot) - Decides how many records from the modern molecular nation flow into the hidden government, so that if wanted, LSTM can handiest select out the long-time period recollections or quick-time period recollections and long-time period recollections.

IV.   Mathematical Formulation

Confusion Matrix is the visual illustration of the particular VS foretold values. It measures the performance of our Machine Learning classification model and appears sort of a table-like structure.

This is. However, a Confusion Matrix of a binary classification downside sounds like

Precision: It may be outlined because of the range of correct outputs provided by the model or, out of all positive categories appropriately foretold by the model, what number of them were valid. It may be calculated as mistreatment by the below formula eq(1)

$$Precision = \frac{TP}{TP+FP} \hspace{3cm} eq(1)$$

Recall: - It is outlined because the out of total positive categories, however our model foretold properly. The recall should be as high as doable

$$Recall = \frac{TP}{TP+FN} \hspace{3cm} eq(2)$$

R factor

$$R_p = \alpha + \beta R_M + \epsilon$$

$R_M = Market\ Return$

$R_p = Portfolio\ Return$

$\epsilon = Error\ Term$

Mean: - In other words, it is by far the most common of the datasets within the diverse fields of arithmetic. As a result, if we take five numbers in a statistics set, say 12, 13, 6, 7, 19, 21, the suggestion system would be

$$\frac{x1 + x2 + x3 + \cdots + xn}{n}$$

Mode: - As a concept, mode refers to the number in a data set that is repetitive and occurs most frequently. The mode is also known as a modal value, which represents the highest number of occurrences in the group. A mode is also a value that represents the whole data collection, like mean and median. There may be more than one mode in a given data set in some cases, so it is imperative to keep this in mind. Bimodal data sets have two modes. As shown in the excel sheet, the mode can be calculated as follows:

Mode.SNGL(B1: B5)

Dataset Used

(https://www.kaggle.com/achintyatripathi/eda-autoviz-class-one-line-code-yahoo-stock-price?scriptVersionId=42446951)

Software requirement specification
- ➢ Python
- ➢ Streamlit web framework

Hardware requirement specification
- ➢ Laptop

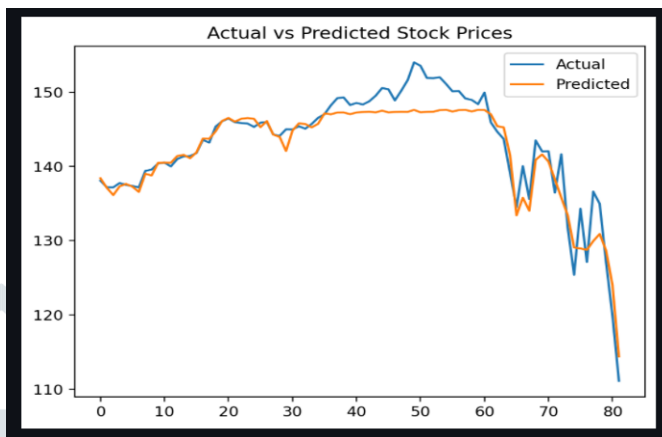V. **RESULTS**

1. New Actual Predicted Result

Figure2. Actual & Predicted Result

Figure2 shows the actual & predicted result.

2. R Factor

Figure3. R factor

Figure3 Shows the R factor result.

The above Fig 3 graph shows the close Price of NKE (Nike) company with actual value denoted in green color and compared with MA (Moving Average), which has MA53 days characterized with blue color and MA30 days is marked with red paint. The graph shows the total daily returns of NKE (Nike) company. On the y-axis, the value is denoted, and on

the x-axis, the year is denoted.

3. Stock Analysis Candle Stick Chart

Figure4. Stock Analysis candle stick chart

In above Fig 4 we see the candle stick chart, which is compared with Boll (Bollinger Band) denoted with red color, RSI (Relative Strength Index) denoted by blue color, and volume by gray color.On the y-axis price of the stock is denoted, and on the x-axis, month and year are denoted.

4. Quantitative Analysis

A quantitative analyst is the only one who designs a complicated framework for monetary establishments that aids them in charging and exchanging securities within the financial market.

Quants may be of types:
Front workplace quant: These are those who immediately offer the dealer the charge of the monetary securities or the buying and selling tools.

Back workplace quant's - These quant are there to validate the framework and create new techniques after research.

Following are the contents of the Quantitative Analysis of NKE (Nike) company.

1. Mean
2. Median
3. Mode
4. Maximum
5. Minimum
6. Variance

Figure5. Quantative Analysis

Figure 6. Quantative Analysis pie chart

In above Figure 6 we see Variance value denoted in blue color, Maximum value in red color, Mean value in green color, Median value in violet color, Mode value in orange color and Minimum value in light blue color.
.

5. Forecasting of Values

Forecasting is used for the time series to predict the value of the stock with the terms of the day like - Short term (5-10 days), medium term (20-100 days), and long term (200 days).NKE stock forecasts on the testing set, testing error 17.41%.The Below Graph Show the predicted value for the previous data.
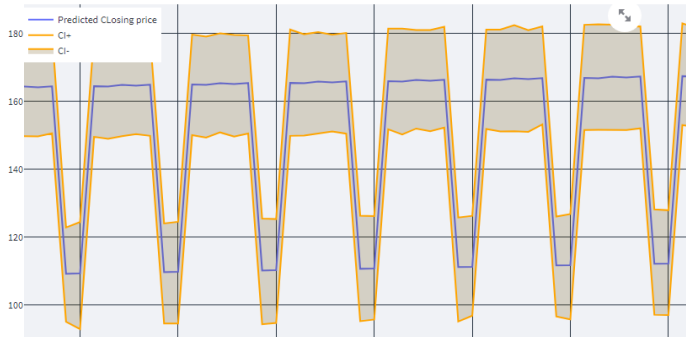The Graph Shows value up to 24 April 2022.

Figure7 forecasting of values

In above Figure 7 Predicted closing price is denoted with a blue color line, CL+ (closing positive) fee, which is above the predicted price and denoted with orange color and also, we see CL- (closing negative) price, which is below the expected price and denoted with orange paint.

6. Confusion Matrix

Figure8 Confusion Matrix

Figure8 shows the confusion matrix.

CONCLUSION

By training with a broader range of knowledge sets, we may be able to improve prediction. It is also possible to analyze specific business aspects within the forecast of varied shares. In this paper, we examine the various patterns of share prices in multiple sectors. The algorithm might analyze a graph with various periods to improve its accuracy. A framework such as this can assist in marketing research and predicting growth for other corporations over several years. Prediction accuracy may be enhanced by incorporating alternative parameters (e.g., capitalist sentiment, election outcome, and government stability).

REFERENCES

1. C Osman Hegazy, Omar S. Soliman, Mustafa Abdul Salam, "A Machine Learning Model for Stock Market Prediction" International Journal of Computer Science and Telecommunications [Volume 4, Issue 12, December 2013].

2. Kai-Yin Woo, Chulin Mai, Michael McAleer, Wing-Keung Wong "Review on Efficiency and Anomalies in Stock Markets" 22 December 2019; Accepted: 4 March 2020; Published: 12 March 2020.

3. Boriss Siliverstovs, Manh Ha Duong "On the role of stock market for real economic activity" JUNE 9 2006.

4. Fahad Almudhaf, Yaser A. AlKulaib OCTOMBER 2012 "ARE CIVETS STOCK MARKETS PREDICTABLE".

5. Hiransha M, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P "NSE Stock Market Prediction Using Deep Learning Models" 2018.

6. Pranav Bhat "A Machine Learning Model for Stock Market Prediction".