JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Stock Market Forecasting using Deep Neural Networks.

Sneha B Masure
JSPM's JSCOE, Pune-28

Dr. S K Hiremath
JSPM's JSCOE, Pune-28

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 retypes 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.

Exponential Smoothing: - When smoothing univariate time series data using the exponential window function, exponential smoothing is a widely used forecasting technique. The technique operates by giving previous data exponentially decreasing weights.

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:

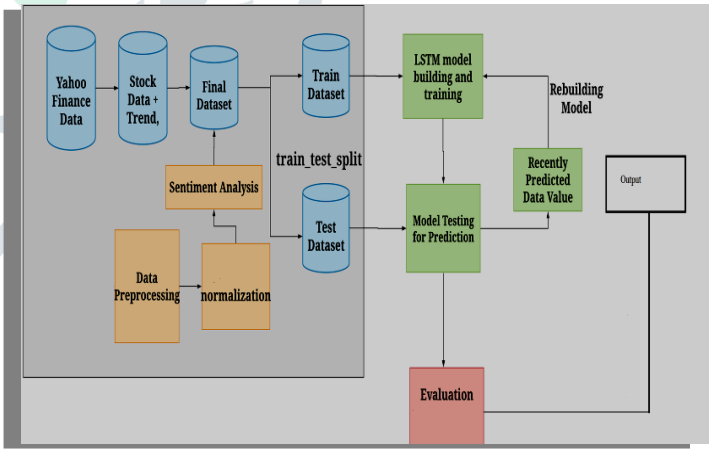

Fig.1. Proposed Architecture

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.

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 handily select out the long-time period recollections or quick-time period recollections and long-time period recollections.

1. **Initialization**: Initialize the LSTM parameters, including weight matrices (W) and bias vectors (b), as well as the initial cell state (C_0) and hidden state (h_0).
2. **For Each Time Step (t)**:
   - **Input (x_t)**: Receive the input for the current time step.
   - **Forget Gate (f_t)**:
     - Calculate the forget gate activation using input x_t and previous hidden state h_{t-1}.
     - Decide what information to forget from the cell state.
   - **Input Gate (i_t)**:
     - Calculate the input gate activation using input x_t and previous hidden state h_{t-1}.
     - Decide what new information to store in the cell state.
   - **Cell State Update (\tilde{C}_t)**:
     - Calculate a new candidate cell state update using input x_t and previous hidden state h_{t-1}.
   - **Cell State (C_t)**:
     - Update the cell state by combining the forget gate, input gate, and candidate cell state update.
   - **Output Gate (o_t)**:
     - Calculate the output gate activation using input x_t and previous hidden state h_{t-1}.
     - Determine what part of the cell state to expose as the hidden state.
   - **Hidden State (h_t)**:
     - Calculate the new hidden state by applying the output gate to the cell state.
   - **Output (output_t)**: (If needed)

- Compute the output for the current time step.

3. **End Loop**: Continue processing for each time step in the sequence.

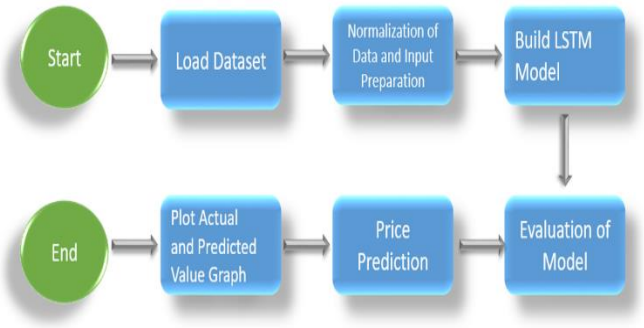

Fig.2 Schematic Diagram for LSTM Module

The figure2 shows the schematic diagram for LSTM module. It consist of essential elements, such as the gates, cell state, and hidden state, are highlighted in this flowchart, which gives a condensed description of how an LSTM cell functions at each time step. Additional information concerning weight matrices, activation functions, and the precise processes required to compute gate activations and cell state updates can be found in practice.

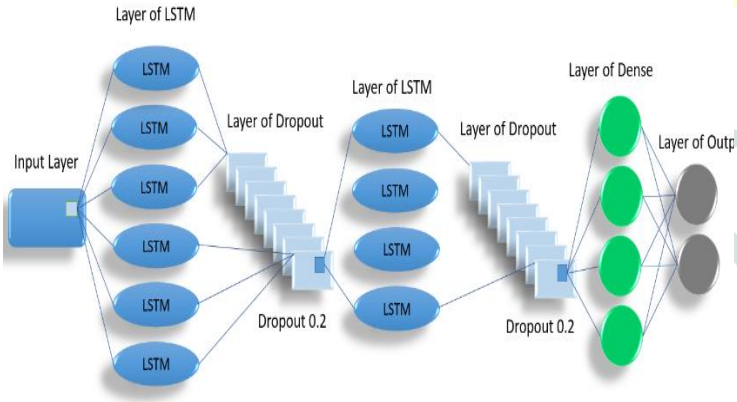

Fig 3 Architecture of LSTM

Random Forest

Famous ML approach used for heart disease prediction due to its exceptional performance and versatility is nothing but RF. It works by creating DT, where each tree is built using an arbitrary amount of training information along with a randomized sample of characteristics from the dataset. This process introduces diversity among the individual trees, reducing the risk of excessive fitting and improving the system's capability for adapting freshly collected information. During training, each decision tree learns to make predictions based on different combinations of features and patterns present in the data. When it comes to making predictions for a new instance (patient), all the individual decision trees in the Random Forestprovide their predictions, and majority vote or an average is used to decide the final forecast. RF‟s strength lies in capacity for receiving complex and connections that aren't linear within the data. It can capture interactions between different various hazard variables, particularly hypertension, and fat levels, smoking status, & more. By combining the predictions from multiple trees, Random Forest achieves high accuracy, making it an effective tool for identifying important risk factors and predicting the likelihood of heart disease in patients. Moreover, Data noise and outliers are resistant to RF, & crucial when dealing with real-worldhealthcare datasets. It also provides a measure of feature importance, indicating which risk factors have the most significant impact on the prediction. This feature importance analysis can help clinicians and researchers better understand the underlying factors contributing to heart disease. Overall, Random Forest's ability to handle complex data, its accuracy, and robustness makes it a valuable and widely used tool for heart disease prediction and risk analysis in healthcare settings. The fig 2 shows flowchart of a RF. The pseudocode of the RF algorithm is as follows

Algorithm: Random Forest

Input: Dataset

Output: Displays the accuracy of RF algorithm

1.Choose k characteristics as arbitrary from a total of m characteristics.

2.Here, k is significantly smaller than m.

3.Using the optimal split point among the k characteristics, determine the node d.

4.Use the best split to divide the node into child nodes.

5.Continue in steps 1 through 3 until the desired number of nodes is achieved.

6.To construct an infinite number of trees, build a forest by repeatedly performing steps 1 through 4.

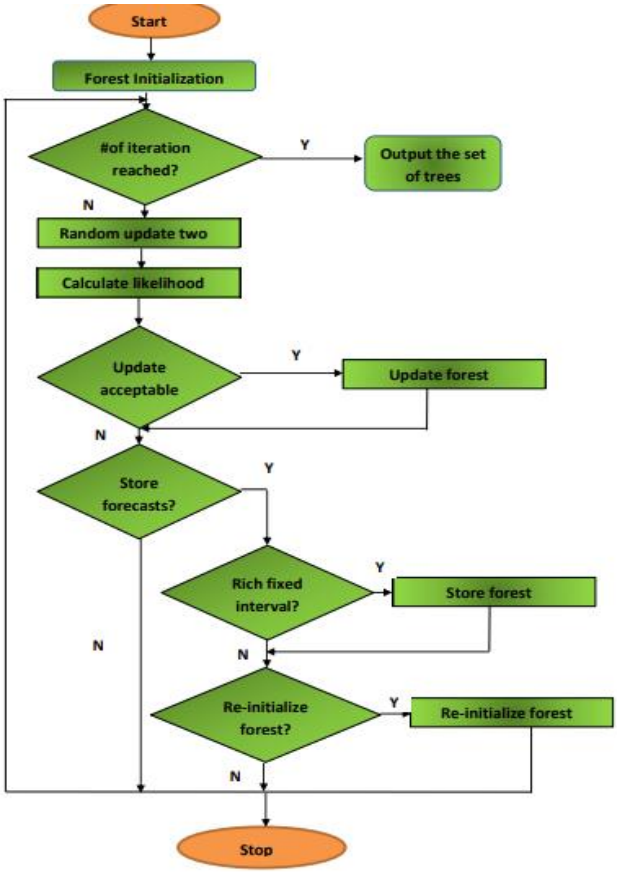

Fig 4 Random Forest Flow chart

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.

A table that summarizes how well a machine learning model performs when put to the test against a dataset is called a confusion matrix. This matrix is frequently used to assess how well classification models which aim to forecast labels for various inputs perform. The matrix displays the counts of different outcomes, including the correct positive predictions (true positives), the correct negative predictions (true negatives), the incorrect positive predictions (false positives), and the incorrect negative predictions (false negatives) made by the model during testing. The matrix is a 2x2 table in binary classification (when there are only two possible classes). The matrix's shape reflects the number of classes in multi-class categorization (when there are more than two classes). For instance, if n classes exist, n x n table will make up the matrix. The actual and anticipated classes are represented, respectively, by the rows and columns of the confusion matrix, which is a square matrix. For a binary classification problem, it has four main components:

1. True Positives (TP): The total volume of instances falling into positive category (class 1) and that the model accurately identified as positive.

2. True Negatives (TN): The total volume of instances falling into negative class (class 0) and that the model accurately identified as positive.

3. False Positives (FP): The amount of cases that fall into the negative category but are incorrectly predicted by model as positive instances is known as Type I mistakes.

4. False Negatives (FN): These instances, commonly referred to as Type II errors, are the amount of positive class incidents that the model mistakenly projected to be negative

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)

$$\text{Precision} = \frac{\text{TP}}{TP+FP} \qquad \text{eq(1)}$$

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

$$\text{Recall} = \frac{\text{TP}}{TP+FN} \qquad \text{eq(2)}$$

R factor

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

R_M = Market Return

R_p = Portfolio Return

ϵ = 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)

V. RESULTS AND Discussion

10 Epochs and 128 LSTM Units

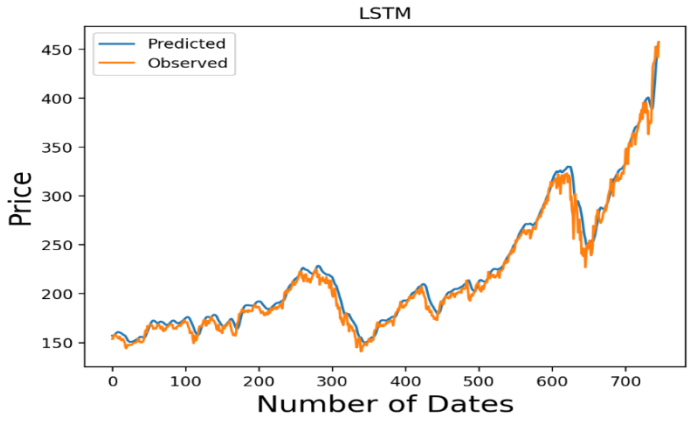

Fig 5 Actual VS Predicted for Amazon Company

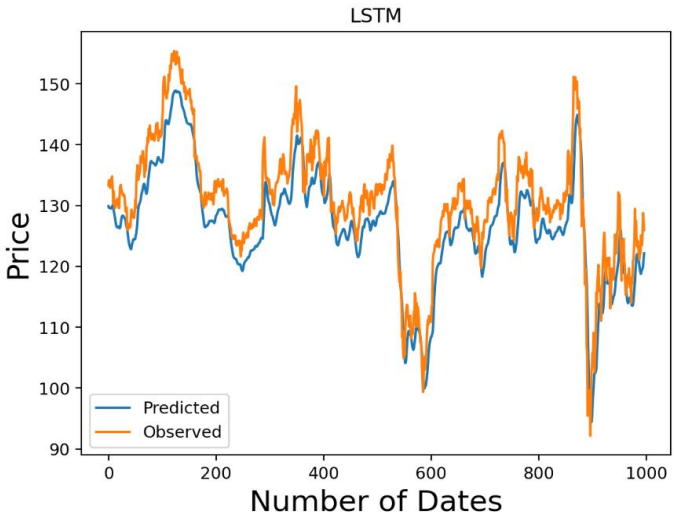

Fig 6 Actual VS Predicted for IBM Company

The figure 5 and figure 6 shows the Actual VS Predicted for Amazon and IBM company respectively The blue line shows the predicted price and orange line shows the observed price.

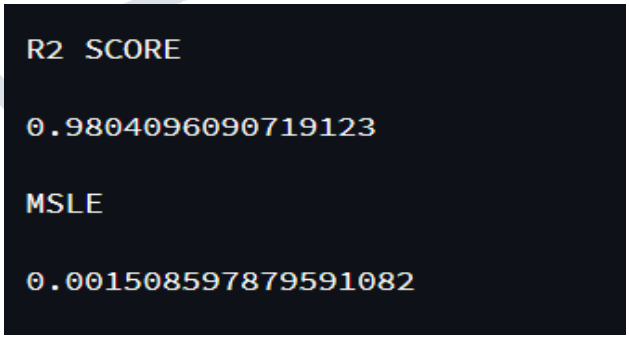

Fig 7 Evaluation Matrix for amazon company

```
R2 SCORE

0.7667341596094519

MSLE

0.0014513421152513 21
```

Fig 7 Evaluation Matrix for IBM company

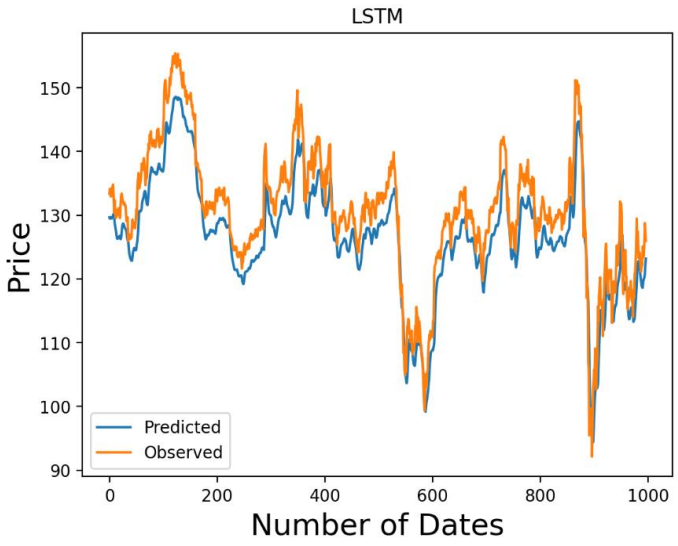

Fig 9 Actual VS Predicted for IBM Company

10 Epochs and 256 LSTM Units

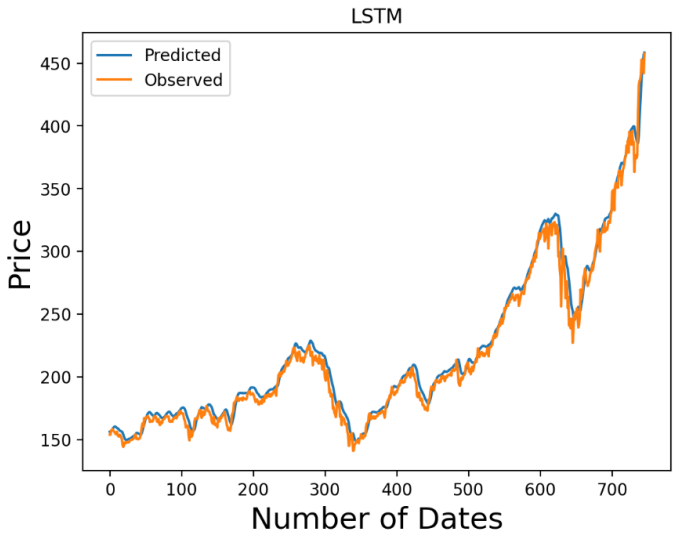

Fig 8 Actual VS Predicted for Amazon Company

```
R2 SCORE

0.9841098198043381

MSLE

0.0011942795114570817
```

Fig. 10 Evaluation Matrix for Amazon Company

```
R2 SCORE

0.7667341596094519

MSLE

0.0014513421152513 21
```

Fig. 11 Evaluation Matrix for IBM Company

15 Epochs and 128 LSTM Units

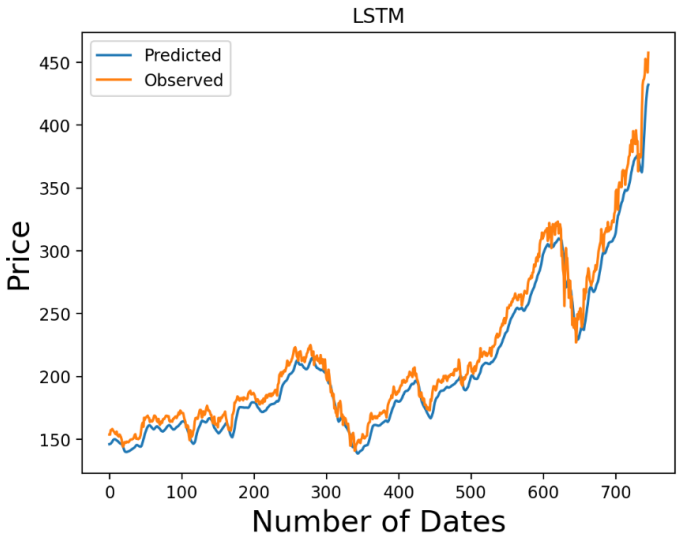

Fig. 12 Actual VS Predicted for Amazon Company

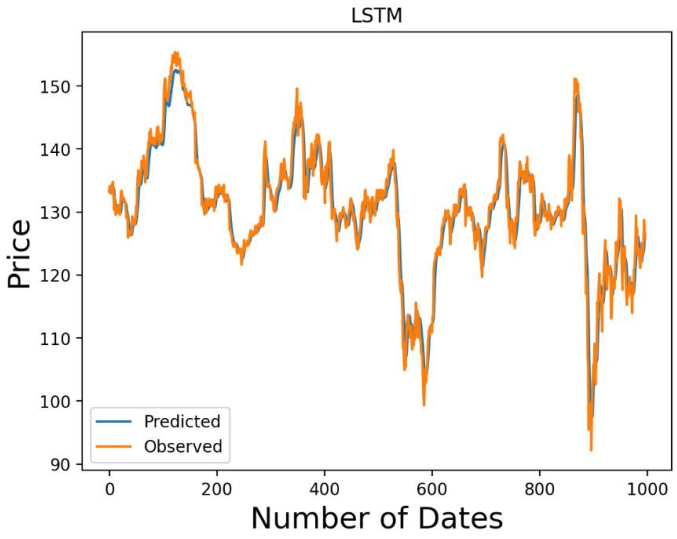

Fig 13 Actual VS Predicted for IBM Company

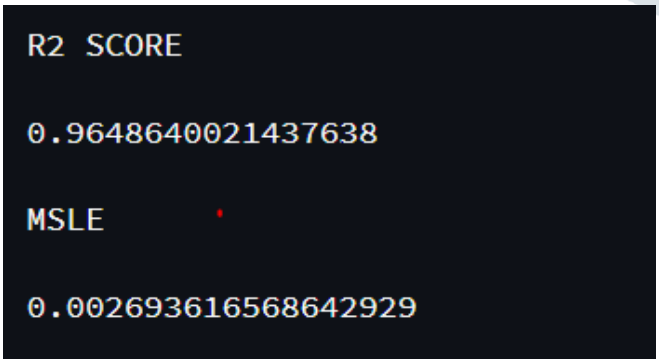

Fig. 14 Evaluation Matrix for Amazon Company

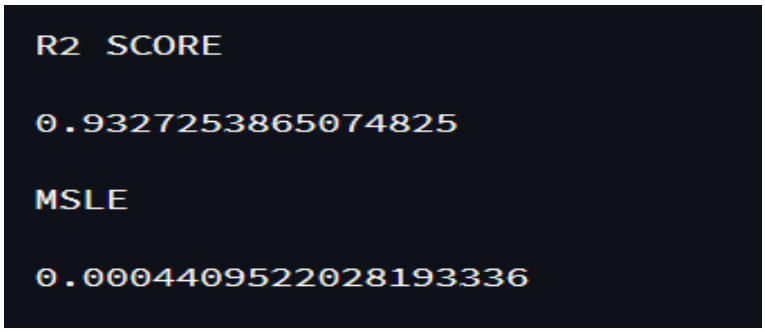

Fig. 15 Evaluation Matrix for IBM Company

Result Analysis

Table1: Values of processing time, R2 Score and MSLE for different epochs and LSTM units

		Apple			IBM		
		Processing Time ms/step	R2 Score	MSLE	Processing Time ms/step	R2 Score	MSLE
10 Epochs	128 LSTM Unit	35ms	0.9804	0.0015	35ms	0.7667	0.0014
	256 LSTM Unit	157ms	0.9841	0.0011	120ms	0.7901	0.0012
15 Epochs	128 LSTM Unit	35ms	0.9648	0.0026	38ms	0.9327	0.0004
	256 LSTM Unit	139ms	0.9809	0.0011	130ms	0.9488	0.0003
25 Epochs	128 LSTM Unit	36ms	0.9716	0.0014	34ms	0.9451	0.0003
	256 LSTM Unit	124ms	0.9707	0.0019	128ms	0.8913	0.0006

CONCLUSION

To enhance prediction by training with a wider variety of knowledge sets. Within the prediction of several shares, it is also feasible to study particular business elements. In this essay, we look at the varied share price trends across a range of industries. To increase its accuracy, the algorithm might examine a network with different eras. This kind of framework can help with marketing analysis and growth forecasting for various

businesses over several years. Alternative parameters (such as capitalist attitude, election results, and government stability) may improve prediction accuracy.

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".