IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204251 |

STOCKVISION: Stock Price Forecasting Using Machine Learning

A. Lydia [1], M. Dinesh [2], P. Somesh [3], R. Jayanth [4], P. Geethika [5]

Assistant Professor, Dept. of CSE., JNTUA University, Kuppam Engineering College, Chittoor, India[1]

UG Students, Dept. of CSE., JNTUA University, Kuppam Engineering College, Chittoor, India[2-5]

ABSTRACT: This project implements a web application using Flask, which predicts the next day's stock prices based on historical stock data. The system leverages deep learning techniques, specifically Long Short-Term Memory (LSTM) networks, to forecast future prices. The workflow involves downloading historical stock data using Yahoo Finance API, preprocessing the data, creating sequences suitable for LSTM training, training the model, and making predictions. The application allows users to input a stock symbol and provides predictions for the next day's closing price, along with high and low price estimates. Additionally, it generates a visual representation comparing true and predicted prices for validation. The system aims to provide users with insights into potential future stock price movements, aiding in investment decision-making.

KEYWORDS: LSTM, CNN, ML, DL, Trade Open, Trade Close, Trade Low, Trade High.

I. INTRODUCTION

Stock market prediction means forecasting the current trends of a company and predict the value of stocks whether it's going up or down. Stock market is the place where a company's shares are traded. A stock is an investment in an institution where it represents ownership in a company. Stock market is a place where those stocks are purchased. Purchasing a stock of a company is owning a small share of an institution.

we are predicting the stock prices using the machine learning algorithm to develop a model which forecasts the stock price effectively based on the current market trends. We have used LSTM recurrent neural networks to predict the stock prices accurately. You would find two types of stocks, one of them was Intraday trading, which is known to us by the term day trading. Intraday trading is that which means all positions are squared-off before the market closes then and there and there would be no possibility of changing the ownership after the day end. LSTM's are very important, as they are very powerful in sequence prediction problems because they could store previous or past information. This is very important in stock prediction as we need to store and read the previous stock information as well to forecast the stock prices accurately in the future.

The rest of the paper is organized as follows. Section 2 introduces the research status of stock price prediction. Section 3 introduces the methodologies. Section 4 consists of the experimental results and the analysis of the results. Section 5 concludes the paper.

II. RELATED WORK

Stock price prediction can be predicted using AI and machine learning models in machine learning fields. Using the SVM model for stock price prediction. SVM is one of the machine learning algorithms 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 prices (Osman Hegazy et al. 2013 [20]; V Kranthi Sai Reddy, 2018 [8];

Random Walk Hypothesis which is proposed by Horne, j. C et al 1997 [27] which is used to predict stock prices, Horne j.c [27] said that the stock values are changes random and the past price values are not dependent on current values. EMH is different from the Random walk hypothesis but the EMH works mainly on Short term patterns for predicting stock prices.

Manh Ha Duong Boris's Siliverstovs, 2006 [11] search the abstraction between equity prices and combined finances in Key Eu nations like UK and Germany. Acceleration in Eu nations investments is apt to results successful even Stronger correlation between the different Eu nations and equity prices. This operation may also lead to a merge in financial

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204251 |

development between EU nations, if advancements in stock markets affect real financial instruments, such as investing and Consuming. Fahad Almudhaf et al, 2012 [22], tests the weak-form market efficiency of CIVETS over the period 2002–2012. The random walk hypothesis process is used in CIVETS. In an efficient stock market, the equity values must follow a random walk hypothesis, when it comes to the future price, the values are changing randomly and unpredictable. Everyday returns for rising and improved markets have been tested for random walks.

LSTM algorithm consists of a Recurrent Neural network to encode data. The algorithm inputs are economic news headings infusion From Bloomberg and Reuters. Long Short- term Memory with embedded layer and the LSTM with the automatic encoder in the stock market for predicting stock values. The Xiongwen Pang et al [4]. Used an automatic encoder and embedded layer to vectorizing the values by using LSTM layers. Correlation coefficients in stocks are selected randomly and predicted using ARIMA and the neural network approach. In this RNN and LSTM algorithms are implemented. M. Nabipour et al [17]. Used different machine learning and deep learning algorithms for predicting stock values such as random forest, decision tree and neural networks. LSTM gives the most accurate results and it has the best ability to fit. LSTM gives the best results while predicting stock prices with the least error rate (Hyeong Kya Choi,2018 [16]; Huicheng Liu, 2018 [15]; M. Nabipour et al,2020 [17]; Xiongwen Pang et al, 2020 [4]).

Recently, Pranab Bhat, 2020 used convolution neural networks for predicting stock values, in this model learning is finished by computing the mean square blunder for each consequent perception and a model is picked that has the least mistake and high prescient power. In this paper, they are utilizing CNN for anticipating stocks and incentives for the following day. Mohammad Mekayel Anik et al, 2020 [23], implemented a linear regression algorithm for future stock price prediction. In this they achieved their goals in predicting accuracy of the model is very good and it might be used for predicting stock values. Xiao Ding et al. 2020 [14] used an easy and effective interface to add common sense knowledge to the process while learning of events.

The LMS filter is a type of adaptive filter which is used for solving linear problems. The idea of the filter is to find the filter coefficients and to minimize a system by reducing the least mean square of the error value (Asep Juarna, 2017 [24]; Eleftherios Giovanis, 2018 [25]). They used a hybrid model for predicting the stock values by using deep learning and ML methodologies and they built a model using deep regression based on CNN. Here they used CNN for parameters, thereby increase the no of loops will stabilize the validation loss. They also tested using DL and a hybrid ML algorithm for stock price prediction. Vivek Rajput and Sarika Bobde [26] used sentiment analysis from online posts or multimedia and data mining is used. In sentiment analysis, they are trying to get emotion either positive or negative based on the textual information available on social networks. sentiment analysis for predicting the stock market to get more accurate and efficient results.

III. PROPOSED ALGORITHM

LSTM uses the RNN approach which has the ability to memorize. Each LSTM cell has three gates i.e. input, forget and output gates. While the data that enters the LSTM's network, the data that is required is kept and the unnecessary data will be forgotten by the forget gate.
LSTM can be used in many applications such as for weather forecasting, NLP, speech recognition, handwriting recognition, time-series prediction, etc.

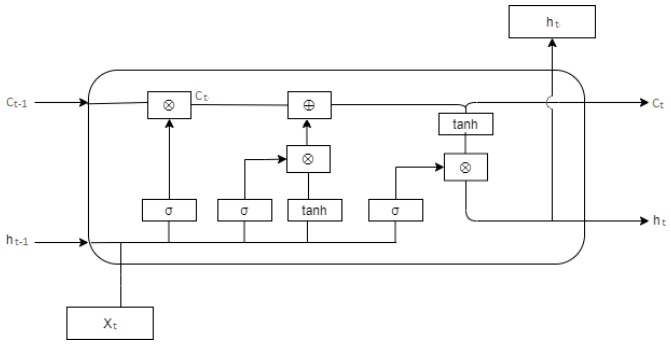

Fig 3.1 LSTM Model

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204251 |

As shown in Fig. 4.1.1, the inputs to the current cell state (C_t) is the previous hidden state (h_{t-1}), previous cell state (C_{t-1}) and present input (X_t). The cell consists of three gates i.e. forget gate, input gate and output gate.

- **Forget Gate:**
  - A forget gate will remove unnecessary data from the cell state.
  - The information that is less important or not required for the LSTM to understand things is removed by performing multiplication of hidden state by a sigmoid function.
  - This step is necessary to optimize the performance of the model.
  - It takes two inputs i.e., $h_{(t-1)}$ and x_t, where $h_{(t-1)}$ is the previous cell hidden state output and x_t is the current cell input.

$$F_t = \sigma\ (W_{fx} * Xt + W_{fh} * h_{t-1} + b_f)$$

- **Input Gate:**
  - This cell is responsible for regulating the data that is added to the cell from the input. Forget gate is used to filter some input.
  - A vector is created by adding all the possible values from the previous cell hidden state $h_{(t-1)}$ and current cell input X_t by using the tanh function. The output of the tanh function in the ranges of [-1, 1].
  - Finally, the outputs of sigmoid and tanh functions are multiplied and the output isadded to the cell state.

$$I_t = \sigma\ (W_{ix} * X_t + W_{hh} * h_{t-1} + b_i) + \tanh(W_{cx} * X_t + W_{ch} * h_{t-1} + b_i)$$

- **Output Gate:**
  - Tanh function is applied to the cell state to create a vector with all possiblevalues.
  - Sigmoid function is applied to previous cell hidden state $h_{(t-1)}$ and current cellinput x_t to filter necessary data from the previous cell.
  - Now, the outputs of sigmoid and tanh functions are multiplied and this output issent as a hidden state of the next cell.

$$O_t = \sigma\ (W_{ox} * X_t + W_{hh} * h_{t-1} + W_{oc} * C_{t-1} + b_i)$$

Intermediate cell state (C_t) is obtained by the multiplication of Forget gate (F_t) with previous cell state (C_{t-1}). Then this intermediate state is added to the output of the input gate.

$$C_t = F_t * C_{t-1} + I_t$$

Current hidden/output state is obtained by multiplying output gate and tanh of cell state.

$$h_t = O_t * \tanh(C_t)$$

3.1 System Architecture:

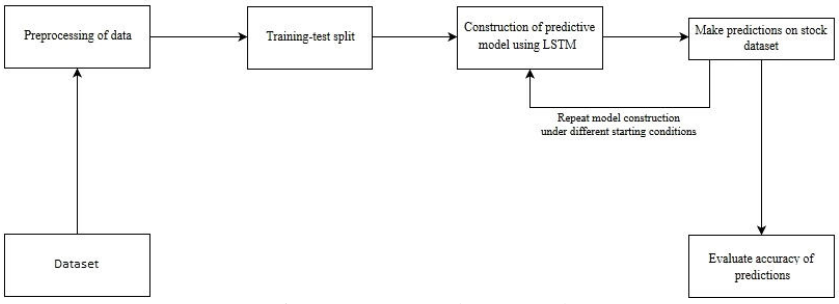

Fig. 3.2. Proposed Approach

Data Selection:
The first step is to select data for an organization and split the data into training and testing. we have used 75% for training and 25% for testing purposes.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204251 |

Pre-processing of data:
In pre-processing, we are selecting attributes required for the algorithm and the remaining attributes are neglected. The selected attributes are Trade Open, Trade High, Trade Low, Trade Close, Trade Volume. In pre-processing, we are using normalization to get values in a particular range.

Prediction using LSTM:
In this system, we are using the LSTM algorithm for predicting stock values. Initially, the training data is passed through the system and train the model.Then in the testing phase, the predicted values are compared with the actual values.

Evaluation:
In the evaluation phase we are calculating the Accuracy, Mean Square Error (MSE) and Root Mean Square Error (RMSE) values for comparison.

IV. SIMULATION RESULTS

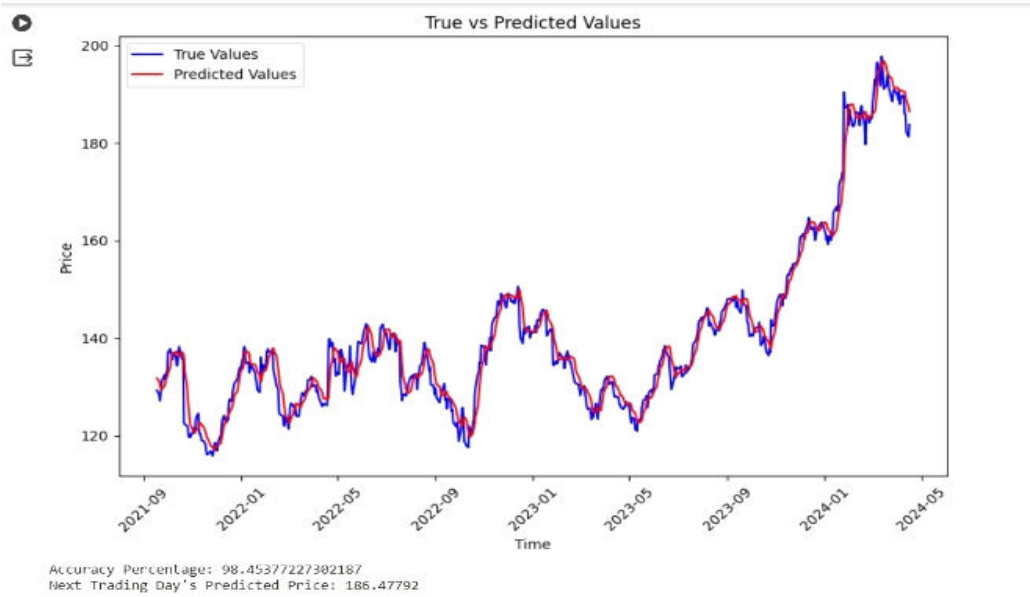

Fig 4.1 Google Stock Original Price and Predicted Price

In the results, as depicted in Fig 4.1, the graph illustrates the Trade Close values for the Google dataset. The graph consists of two lines: the blue line represents the actual values observed during training, while the yellow line represents the predicted values generated from the test data. This visual representation allows for a direct comparison between the actual and predicted values of the Trade Close prices over the specified time period.

Additionally, Table 4.1 presents the accuracy metrics, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) values for various iterations (epochs) of the training process. These metrics offer quantitative insights into the performance of the LSTM model across different training iterations. By analyzing these values, we can assess the model's convergence and determine the optimal number of epochs needed to achieve the desired prediction accuracy.

epochs	Accuracy	MSE	RMSE
10	93.00717	207.6578	14.41034
20	94.01166	156.3873	12.50549
30	95.64188	105.3248	10.26279
40	95.59026	99.17409	9.958619
50	96.99466	62.24641	7.88964

Table 4.1 Epochs

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204251 |

In this table, the different epochs represent the number of iterations or passes over the entire training dataset during the training process of the LSTM model. Each row corresponds to a specific epoch, and the corresponding values in the columns represent the following:

Epochs: The number of training epochs, indicating how many times the model has been trained on the entire dataset.

Accuracy (%): The accuracy of the LSTM model in predicting stock prices, expressed as a percentage. It represents the proportion of correct predictions made by the model compared to the total number of predictions.

Mean Squared Error (MSE): The average of the squared differences between the predicted and actual stock prices. It quantifies the average squared deviation of the predicted values from the actual values, with lower values indicating better model performance.

Root Mean Squared Error (RMSE): The square root of the mean squared error, representing the standard deviation of the prediction errors. It provides a measure of the spread of prediction errors and is interpreted in the same unit as the target variable (stock prices).

V. CONCLUSION AND FUTURE WORK

In this project, we successfully developed a web application for stock price prediction using LSTM neural networks. The application allows users to input a stock symbol and receive predictions for the next trading day's closing price, along with the predicted low and high prices. Through this project, we achieved the following key objectives:

- Utilized historical stock price data obtained from Yahoo Finance, preprocessed it, and prepared it for model training.
- Designed and implemented an LSTM neural network architecture capable of capturing temporal dependencies in sequential data.
- Trained the model using a portion of the data and evaluated its performance on a separate test set.
- Provided users with insightful predictions to support their investment decisions.

REFERENCES

1. Mehtab, S. and Sen, J.: A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing. In: Proceedings of the 7th Int. Conf. on Business Analytics and Intelligence, Bangalore, India, December 5 – 7 (2019)
2. Jaffe, J., Keim, D. B. Keim, and Westerfield, R.: Earnings Yields, Market Values and Stock Returns. Journal of Finance, 44, 135 - 148 (1989)
3. Fama, E. F. and French, K. R.: Size and Book-to-Market Factors in Earnings and Returns. Journal of Finance, 50(1), 131 - 155 (1995)
4. Chui, A. and Wei, K.: Book-to-Market, Firm Size, and the Turn of the Year Effect: Evidence from Pacific Basin Emerging Markets. Pacific-Basin Finance Journal, 6(3-4), 275 – 293 (1998)
5. Basu, S.: The Relationship between Earnings Yield, Market Value and Return for NYSE Common Stocks: Further Evidence. Journal of Financial Economics, 12(1), 129 – 156 (1983)
6. Rosenberg, B., Reid, K., Lanstein, R.: Persuasive Evidence of Market Inefficiency. Journal of Portfolio Management, 11, 9 – 17 (1985)
7. Adebiyi, A., Adewumi, O., and Ayo, C. K. Ayo: Stock Price Prediction Using the ARIMA Model. In: Proc. of the Int. Conf. on Computer Modelling and Simulation, Cambridge, UK, pp. 105 - 111 (2014)
8. Mondal, P., Shit, L., and Goswami, S.: Study of Effectiveness of Time Series Modeling (ARMA) in Forecasting Stock Prices. Int. Journal of Computer Science, Engineering and Applications, 4, 13 - 29 (2014)
9. Jarrett, J. E. and Kyper, E.: ARIMA Modelling with Intervention to Forecast and Analyze Chinese Stock Prices. Int. Journal of Engineering Business Management, 3(3), 53 – 58 (2011)
10. Mishra, S.: The Quantile Regression Approach to Analysis of Dynamic Interaction between Exchange Rate and Stock Returns in Emerging Markets: Case of BRIC Nations. IUP Journal of Financial Risk Management, 13(1), 7 – 27 (2016)
11. Mostafa, M.: Forecasting Stock Exchange Movements Using Neural Networks: Empirical Evidence from Kuwait. Expert Systems with Applications, 37, 6302 - 6309 (2010)
12. Wu, Q., Chen, Y., and Liu, Z.: Ensemble Model of Intelligent Paradigms for Stock Market Forecasting. In: Proc. of the IEEE 1st Int. Workshop on Knowledge Discovery and Data Mining, Washington DC, USA, pp. 205 – 208 (2008)
13. Siddiqui, T. A. and Abdullah, Y.: Developing a Nonlinear Model to Predict Stock Prices in India: An Artificial Neural Networks Approach. IUP Journal of Applied Finance,21(3), 36 – 39 (2015)
14. Jaruszewicz, M. and Mandziuk, J.: One Day Prediction of Nikkei Index Considering Information from Other Stock Markets. In: Proc. of the Int. Conf. on Artificial Intelligence and Soft Computing, Tokyo, pp. 1130 – 1135 (2004)

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details