International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at : www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET2411111

A Novel Approach to predict the Stock Price using LSTM and Linear Regression

Preetjot Kaur, Karan Marwaha, Keshav Kumar

Department of CSE, Dr. Akhilesh Das Gupta Institute of Professional Studies, Delhi, India

ARTICLEINFO

Article History :

Accepted: 02 Jan 2024
Published: 11 Jan 2024

Publication Issue :
Volume 11, Issue 1
January-February-2024
Page Number :
118-123

ABSTRACT

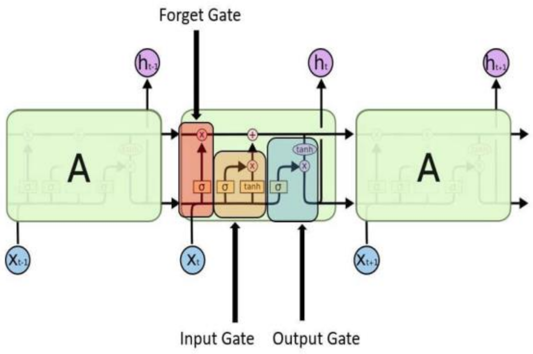
Stock price prediction is a challenging and crucial task in financial markets. Traditional methods often struggle to capture the complex patterns present in stock price movements. In this study, we propose a hybrid model combining Long Short-Term Memory (LSTM) and Linear Regression techniques to improve the accuracy and robustness of stock price predictions. We evaluate the performance of our hybrid model using historical stock price data and compare it with individual LSTM and linear regression models. The experiments demonstrate that the hybrid model outperforms the standalone models in terms of accuracy and robustness.

Keywords : Stock Price Prediction, Machine Learning, Linear Regression, long short-term memory

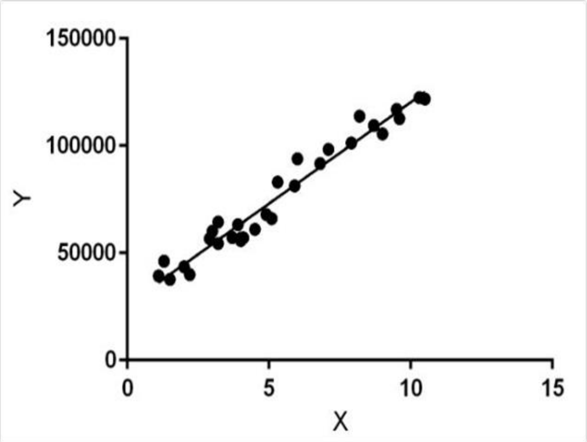
I. INTRODUCTION

The prediction of stock prices has been a due to

the dynamic and often unpredictable nature of financial markets. longstanding challenge in the field of financial forecasting Investors, traders, and financial analysts continuously seek accurate and reliable methods to anticipate future stock price movements for informed decision-making. In recent years, machine learning techniques, particularly deep learning models, have gained popularity for their ability to capture complex patterns in time-series data. This study focuses on the application of two distinct yet complementary techniques, Long Short-Term Memory (LSTM) and Linear Regression, to enhance the accuracy of stock price predictions. LSTM, a type of recurrent neural network (RNN), excels at capturing long-term dependencies and non-linear patterns in sequential data. In RNN output from the last step is fed as input in the current step. RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from recent information. As the gap length increases RNN does not give an efficient performance. Long Short-Term Memory (LSTM) is specifically designed to handle sequential data, such as time series. It tackled the problem of long-term dependencies of RNN. LSTMs address this problem by introducing a memory cell, which is a container that can hold information for an extended period. The memory cell is controlled by three gates: the input gate, the forget gate, and the output gate. These gates decide what information to add to, remove from, and output from the memory cell.



On the other hand, Linear Regression, a classical statistical method, is adapted to modeling linear relationships and capturing short-term fluctuations in data. It is also a type of machine-learning algorithm more specifically a supervised machine-learning algorithm that learns from the labeled datasets and maps the data points to the most optimized linear functions. which can be used for prediction on new datasets. The goal of the algorithm is to find the best linear equation that can predict the value of the dependent variable based on the independent variables. The equation provides a straight line that represents the relationship between the dependent and independent variables. The slope of the line indicates how much the dependent variable changes for a unit change in the independent variable(s). Here Y is called a dependent or target variable and X is called an independent variable also known as the predictor of X = Date and Y = price



The integration of LSTM and Linear Regression in a hybrid model aims to leverage the strengths of both approaches, providing a more robust and accurate prediction tool for stock prices. The LSTM network acts as a sophisticated feature extractor, learning intricate patterns and dependencies from historical stock price data. Simultaneously, the Linear Regression component complements the LSTM by capturing linear trends and adjusting for short-term variations, contributing to a more comprehensive understanding of market dynamics. The significance of this research lies in its potential to offer a holistic solution that combines the power of deep learning with the interpretability of traditional statistical methods. By integrating LSTM and Linear Regression, the proposed model seeks to overcome the limitations of individual techniques and provide a more resilient and adaptable approach to stock price prediction. The subsequent sections of this study will delve into the methodology, data sources, experimental setup, and results to comprehensively evaluate the effectiveness of the hybrid LSTM-Linear Regression model in the context of stock market forecasting.

II. RELATED WORK

Numerous studies have explored the effectiveness of LSTM networks in predicting stock prices. Gao et al. (2017) applied LSTMs to model the temporal dependencies in financial time series data, demonstrating improved predictive performance compared to traditional methods.

Chen et al. (2019) used a hybrid approach combining linear regression and feature engineering techniques. Feature engineering involved the extraction of financial indicators and sentiment analysis from news articles, providing additional contextual information for the linear regression model.

Ensemble models combining LSTM and linear regression have also been investigated. Zheng et al. (2021) proposed an ensemble model that dynamically combines predictions from LSTM and linear regression based on the current market conditions, achieving improved overall accuracy.

In this paper, we took the hybrid LSTM-Linear Regression model proposed which aims to contribute to this body of knowledge by offering a synergistic approach that leverages the advantages of both deep learning and traditional statistical methods for enhanced predictive accuracy.

III. DATA COLLECTION AND PREPROCESSING

A. Data Collection

We extracted the total values of AAPL from January 1st, 2010 to December 15th, 2023. Apple has been a leading technology company, and its revenue and net income have experienced consistent growth over the years. As of my last update in 2022, Apple's annual revenue was in the hundreds of billions of U.S. dollars. Its primary sources of revenue include the sales of iPhones, iPads, Macs, wearables, and services such as the App Store, iCloud, and Apple Music. Apple's stock, traded under the ticker symbol AAPL, has historically been one of the most valuable and widely held stocks. Stock opening and closing prices of AAPL from January 1st, 2010 to December 15th, 2023 are obtained from Yahoo! Finance.

B. Data Pre-Processing

Stock prices data collected is not complete understandably because of weekends and public holidays when the stock market does not function. The missing data is approximated using a simple technique. Stock data usually follows a concave function. So, if the stock value on a day is x and the next value present is y with some missing in between. The first missing value is approximated to be (y+x)/2 and the same method is followed to fill all the gaps. We split the stock data dataset into training, validation, and testing sets, adhering to chronological order to simulate real-world use cases.

IV. METHODOLOGY

Prediction is the main task where we predict the value of the stock for that we gather historical data which is preprocessed, cleaned, and formatted appropriately. This includes handling missing data, normalizing numerical features, and creating sequences for LSTM.

A. Model training:

We prepared the training data for a time series prediction model, particularly for LSTM (Long Short-Term Memory) networks. Additionally, We set up the data for linear regression.

Splitting our preprocessed data into training and testing sets. This is crucial to evaluate the model's performance on unseen data.

	Open	High	Low	Close	Volume
0	7.62250	7.660714	7.585000	7.643214	493729600
1	7.664286	7.699643	7.616071	7.656429	601904800
2	7.656429	7.686786	7.526786	7.534643	552160000
3	7.562500	7.571429	7.466071	7.520714	477131200
4	7.510714	7.571429	7.466429	7.570714	447610800

Training data is the DataFrame that contains the first 70% of the 'Close' column data, assumed to be the training set. While, Testing data is the DataFrame that contains the remaining 30% of the 'Close' column data, assumed to be the testing set. We used MinMax scaling from Sklearn.preprocessing to scale the data to a specified range, in this case, between 0 and 1, and used fit.transform that fits the MinMaxScaler to the training data and transforms it. The 'Close' prices are scaled to the range [0, 1].

B. Creating Sequence for LSTM

We created a loop that creates sequences for training an LSTM model. It takes the last 100 consecutive 'Close' prices as input (X_train) and the next 'Close' price as the corresponding output (Y_train). Where X_train is a 3D array where each element is a sequence of 100 'Close' prices and Y_train is a 1D array representing the target values (next 'Close' prices). The generated X_train and Y_train arrays are used to train an LSTM model for stock price prediction.

C. Sequential Model

Sequential creates a linear stack of layers for building the neural network. This architecture consists of multiple LSTM layers with increasing units and dropout rates, which is a common approach to capturing complex patterns in time series data. The final dense layer outputs a single value, representing the predicted stock price. We use the pd.concat method that concatenates two Dataframes along a particular axis. In this case, it concatenates part_100_days and data_testing along the rows (axis=0) and we use a parameter that is set to True to reset the index of the resulting DataFrame.

Model: "sequential"

LAYER(type)	OUTPUT SHAPE	PARAM
LSTM	(None,100,50)	10400
Dropout	(None,100,50)	0
LSTM 1	(None,100,60)	26640
Dropout 1	(None,100,60)	0
LSTM 2	(None,100,80)	45120
Dropout 2	(None,100,80)	0
LSTM 3	(None,120)	96480
Dropout 3	(None,120)	0
Dense	(None,1)	121

We repeat the same process for testing data and iterate it over a loop starting from the 100th index (i=100) to the end of the data. For each iteration, it takes the previous 100 data points (input_data [i-100: i]) as a sequence and appends it to X_test. This creates sequences of 100-time steps. The target values (Y_test) are the next data point in the sequence, specifically input_data[i, 0]. It assumes that the target variable you want to predict is in the first column of your data. Finally, the lists X_test and Y_test are converted to NumPy arrays. Where, X_test will contain the sequences of 100-time steps, and Y_test will contain the corresponding target values. We use a model. predict to predict the target values of input test data and use the scaler_factor to It is used to reverse the scaling transformation.

After this, Y_predicted and Y_test should represent the predicted and actual target values in their original scale.

V. RESULTS AND DISCUSSION

This section gives an overview of the LSTM-Linear regression model.

Collected data from Yahoo finance of AAPL stock from 2010 to 2023 and plotted the closing price concerning time on the graph for our reference to the stock's actual price.

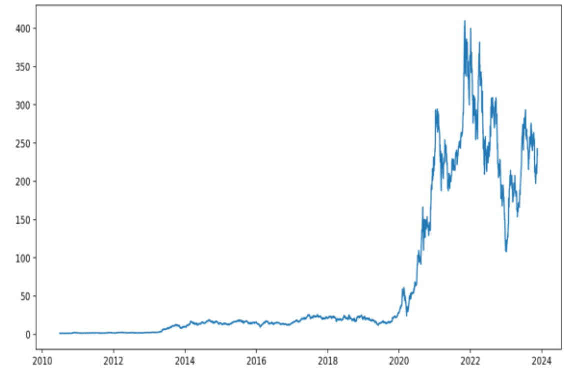


Figure 1 Closing Price vs Time of AAPL

100 moving average of the data concerning the actual price. The 100-day moving average helps smooth out short-term fluctuations and noise in the daily stock price movements. This can make it easier to identify the underlying trend by providing a clearer view of the overall direction in which the stock price is moving. It often uses the MA100 to identify the prevailing trend. If the current stock price is above the MA100, it might be considered an indication of an upward trend. Conversely, if the stock price is below the MA100, it might signal a potential downward trend.

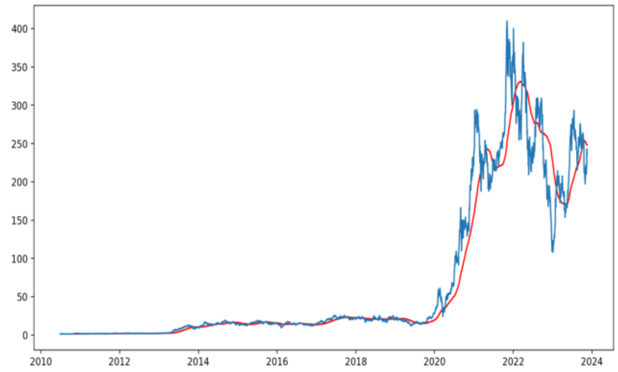


Figure 2 Closing Price vs Time with 100MA

While the 200-day moving average is used for a longer-term perspective. It helps identify the

broader, more sustained trends in the stock price. The MA200 is often considered more significant than the MA100 for confirming the overall trend. A stock trading above the MA200 might be considered in a long-term uptrend, while one trading below it might be considered in a long-term downtrend.

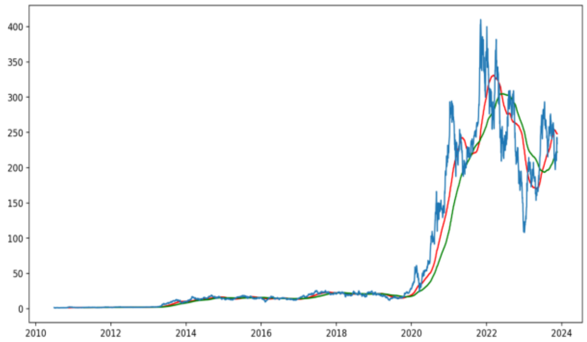


Figure 2 Closing Price vs Time with 100MA & 200MA

Here the blue line indicates the closing price, the red line indicates the 100 moving average of the closing price, while the green line indicates the 200 moving average of the closing price which gives more detailed explanation about the stock trends.

Actual price and predicted closing price of AAPL, a large stock. The model was trained in bulk sizes of 512 and 50 epochs, and the forecasts were made very similar to stock prices, as seen in the graph.

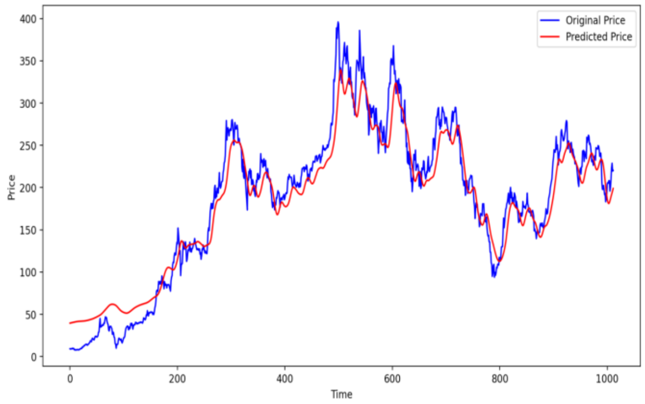


Figure 4 Predicted stock price for AAPL

VI. CONCLUSION

The integration of Long Short-Term Memory (LSTM) and linear regression models in the domain of stock price prediction presents a robust and versatile approach. By combining LSTM and linear regression, we leverage the strengths of both models. The LSTM captures nuanced patterns, while linear regression enhances interpretability. The ensemble of these models aims to strike a balance between complexity and transparency, providing valuable insights into stock price movements.

In conclusion, the fusion of LSTM and linear regression models offers a promising avenue for advancing the field of stock price prediction. As financial markets evolve, embracing innovative and adaptive approaches becomes imperative. This research serves as a stepping stone, inviting further exploration and refinement of models to meet the ever-changing demands of the financial landscape.

VII. REFERENCES

1. Conrad, J., & Kaul, G. (1998). An anatomy of trading strategies. Review of Financial Studies, 11, 489–515
2. Roondiwala, Murtaza, Harshal Patel, and Shraddha Varma. "Predicting stock prices using LSTM." International Journal of Science and Research (IJSR) 6.4 (2017): 1754-1756.
3. Dai, Z., Zhou, H., Wen, F., & He, S. (2020a). Efficient predictability of stock return volatility: The role of stock market implied volatility. The North American Journal of Economics and Finance, 52, 101174.
4. Dangl, T., & Halling, M. (2012). Predictive regressions with time-varying coefficients. Journal of Financial Economics, 106,
5. Masoud, Najeb MH. (2017) “The impact of stock market performance upon economic growth.” International Journal of Economics and Financial Issues 3 (4) : 788–798.
6. Mr. Amit B. Suthar, Ms. Hiral R. Patel, Dr. Satyen M. Parikh, “A Comparative Study on Financial Stock Market Prediction Models”, 2012.
7. Vinod Mehta at el. , “Stock Price Prediction Using Regression And Artificial Neural Network”, 2017.
8. Ryo Akita, Akira Yoshihara, Takashi Matsubara, Kuniaki Uehara, “Deep learning for stock prediction using numerical and textual information”, 2016.
9. Oyeyemi, Elijah O., Lee-Anne McKinnell, and Allon WV Poole. (2007) “Neural network-based prediction techniques for global modeling of M (3000) F2 ionospheric parameter.” Advances in Space Research 39 (5) : 643-650.
10. Zhang, G. Peter. (2003) “Time series forecasting using a hybrid ARIMA and neural network mode.” Neurocomputing 50 : 159-175.

Cite this article as :

Preetjot Kaur, Karan Marwaha, Keshav Kumar, "A Novel Approach to predict the Stock Price using LSTM and Linear Regression", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 11 Issue 1, pp. 118-123, January-February 2024. Available at doi : https://doi.org/10.32628/IJSRSET2411111
Journal URL : https://ijsrset.com/IJSRSET2411111