Stock Price Prediction using Machine learning

[2]School of Computer Science and Engineering, VIT-AP University, Near Vijayawada, 522237, Andhra Pradesh, India

Narendar Reddy Pathakuntla
narendarreddypathakuntla@gmail.com

Abstr[2]act

For financial analysts and investors to make well-informed decisions, accurate stock price forecasting is crucial. Machine learning techniques have shown to be useful tools for stock price forecasting due to the growing complexity of financial markets and the impact of multiple factors. The efficacy of several machine learning models, such as LSTM, KNN, Neural Network Regression, Light Gradient Boosting Machine, and XGBoost, is investigated in this study using historical stock data and market indicators including trade volume, open, high, low, and close. The effectiveness of these models is evaluated using metrics like Mean Squared Error and R-squared. With an R-squared value of 0.999702 and an MSE of 5.125808, the findings show that the XGBoost model performs exceptionally well in terms of accuracy.

Keywords: Stock Price, Light Gradient boosting machine, Neuron Network Regression, XGBoost, KNN

1 Introduction

Predicting stock prices is essential for traders and investors to make wise choices in the financial markets. Apple Inc. is one of the most significant firms among the many stocks that are traded; investors all over the world own and closely watch its stock. Accurately predicting the changes of Apple's stock price is of great interest to analysts, investors, and financial institutions. Recent developments in machine learning (ML) algorithms have transformed the study and forecasting of the stock market. A key component of stock price forecasting is the use of ML algorithms. In order to help investors make wise decisions and maximize returns, this study focuses on using ML algorithms to anticipate Apple's stock prices. LSTM, KNN, Neural

Network Regression, Light Gradient Boosting Machine, and XGBoost are among the machine learning algorithms that are investigated for their ability to predict Apple's stock values from past data.Based on the aforementioned techniques, the results indicate that the XG Boost model provided the highest level of accuracy.

2 Stock Price

It is vital to examine a multitude of factors that impact the stock market in order to forecast the price of Apple stock. To make informed decisions, investors must be able to understand stock price movements. A multitude of factors, such as company performance, economic indicators, market sentiment, and external events, can impact stock values. For instance, new introductions, quarterly earnings reports, and modifications to industry laws can all affect investor confidence and result in fluctuations in stock values. Moreover, market movements and investor behavior have a significant influence on changes in stock values. Factors like trade volume and open, high, low, and close data are essential for stock price prediction.

Machine learning algorithms can be used to examine these diverse data sets and produce prediction models that help investors anticipate future price changes. Machine learning algorithms that use historical stock data and relevant market indicators can provide investors with insights into potential price movements. As a result, traders are able to make wise choices.

Table 1. Minimum, maximum, and standard deviation of the parameters

Parameters	Minimum	Maximum	Standard Deviation
High	11.1328	705.0708	132.178818
Low	10.9984	699.5688	129.775714
Open	11.1328	702.4108	131.053038
Close	10.9984	702.1	131.017984

3 Machine Learning Techniques

The previous few decades have seen significant advancements in computer technology, which has led to an increase in the collection of electronic data in most fields of human effort. Numerous businesses require enormous amounts of data that go back many years. This data includes personal information, financial activity, biological information, etc. Concurrently, data scientists have been developing algorithms, which are iterative computer programs that have the capacity to examine enormous volumes of data, assess it, and identify patterns and connections that humans are unable to. Studying past occurrences can provide a plethora of knowledge on what to anticipate from similar or nearly similar events in the future.

These algorithms could be able to draw lessons from the past and apply those lessons to future decision-making. The concept of data analysis is not new. Machine learning algorithms set themselves apart from other methods by their ability to handle massive volumes of data and data with little organization. It makes ML algorithms useful in a variety of applications that were previously deemed to be too complex for traditional learning methods.

Five machine learning algorithms—LSTM, KNN, Neural Network Regression, Light Gradient Boosting Machine, and XGBoost—were developed and used in the current work to forecast future stock prices.

3.1 Long Short-Term Memory (LSTM):

LSTM is a type of recurrent neural network (RNN) designed to capture long-term dependencies in sequential data, making it particularly useful for time-series forecasting tasks such as stock price prediction. Unlike traditional feedforward neural networks, LSTM networks have a mechanism called a "memory cell" that allows them to remember information over long periods. This memory cell consists of three gates: the input gate, the forget gate, and the output gate. These gates regulate the flow of information into and out of the cell, allowing LSTM networks to selectively retain or discard information from previous time steps.

3.2 K-Nearest Neighbors (KNN):

KNN is a simple, non-parametric algorithm used for classification and regression tasks. In KNN regression, the predicted value for a data point is the average of the values of its k nearest neighbors in the feature space. The algorithm operates based on the assumption that similar data points tend to have similar target values. The key steps of the KNN algorithm are:

Calculate the distance between the query point and all data points in the training set using a distance metric such as Euclidean distance.Select the k nearest neighbors based on the calculated distances.For regression, predict the target value of the query point as the average of the target values of its nearest neighbors.

3.3 Neural Network Regression:

Neural network regression involves training a neural network to learn a mapping from input features to continuous target values. The network comprises multiple layers of interconnected neurons, including an input layer, one or more hidden layers, and an output layer. Each neuron applies a weighted sum of its inputs, followed by a non-linear activation function to introduce non-linearity into the model. The training process involves adjusting the weights of the connections between neurons to minimize a loss function, typically the mean squared error between the predicted and actual target values.

3.4 Light Gradient Boosting Machine (LightGBM):

LightGBM is a gradient boosting framework that uses decision trees as base learners. Unlike traditional gradient boosting methods that grow trees sequentially, LightGBM uses a technique called "leaf-wise" tree growth, where it splits nodes in a depth-wise manner based on the leaf-wise criterion. This approach results in faster training times and lower memory usage. LightGBM also employs a gradient-based optimization algorithm to grow trees, allowing it to handle large-scale datasets efficiently. Additionally, LightGBM supports customizable objective functions and evaluation metrics, making it versatile for various regression tasks.

3.5 XGBoost:

XGBoost, short for Extreme Gradient Boosting, is another gradient boosting library known for its scalability and performance. Similar to LightGBM, XGBoost builds an ensemble of weak learners (decision trees) sequentially, with each subsequent tree correcting the errors made by the previous ones. XGBoost incorporates several regularization techniques to prevent overfitting, such as shrinkage (learning rate) and tree pruning. It also utilizes parallel processing and cache optimization to enhance training speed. The objective function optimized by XGBoost combines a loss term and a regularization term, and the trees are grown level-wise rather than leaf-wise, striking a balance between computational efficiency and model accuracy.

3.6 Evalution Metrics

The performance of developed algorithms are measured by mean sqaure error (MSE), and coefficient of determination (R^2). The equations are as follows [7] :

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

4 Results and Discussions

Careful data processing and collecting are essential to the construction of a strong stock market price prediction model since they provide the basis for precise forecasts. As a first step, we assembled large datasets that included a range of financial metrics that were essential for developing our prediction models. Our dataset includes historical stock prices and trade volume as important factors. Our prediction model's effectiveness was assessed using a variety of sophisticated machine learning methods and algorithms. In particular, we used LightGBM, K-Nearest Neighbors (KNN), XGBoost, Long Short-Term Memory (LSTM), and Neural Network Regression in our investigation. We used popular libraries like Pandas, scikit-learn, NumPy, and Matplotlib to build these algorithms using Python in the Jupyter notebook.

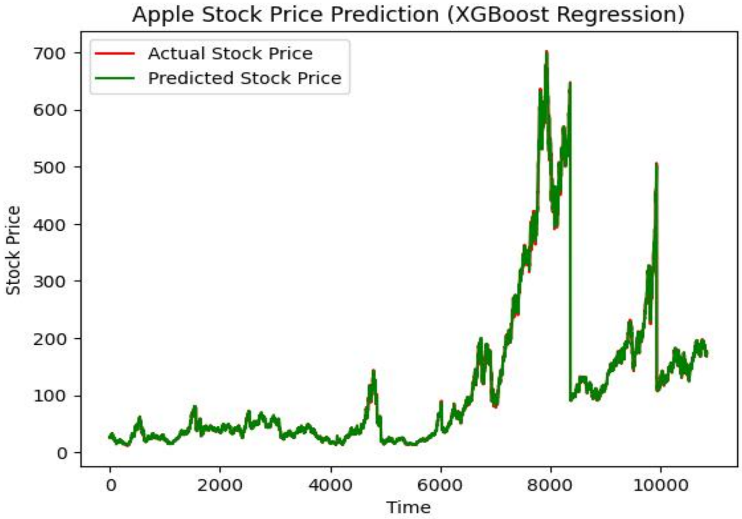


Fig. 1. Plot of Actual Stock Price vs Predicted Stock Price of XGBoost

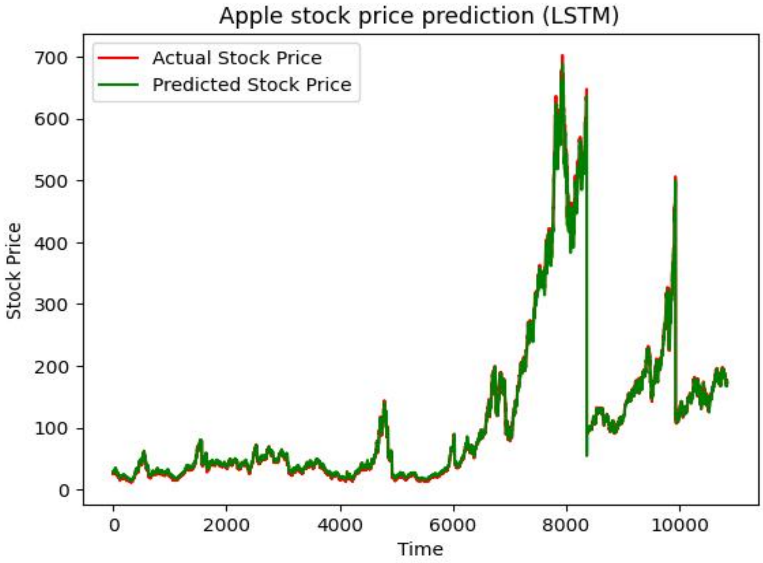
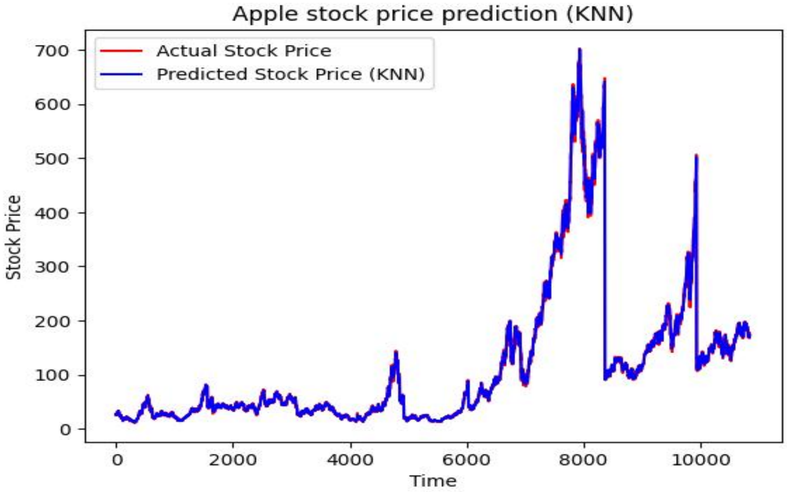


Fig. 2. Plot of Actual Stock Price vs Predicted Stock Price of LSTM



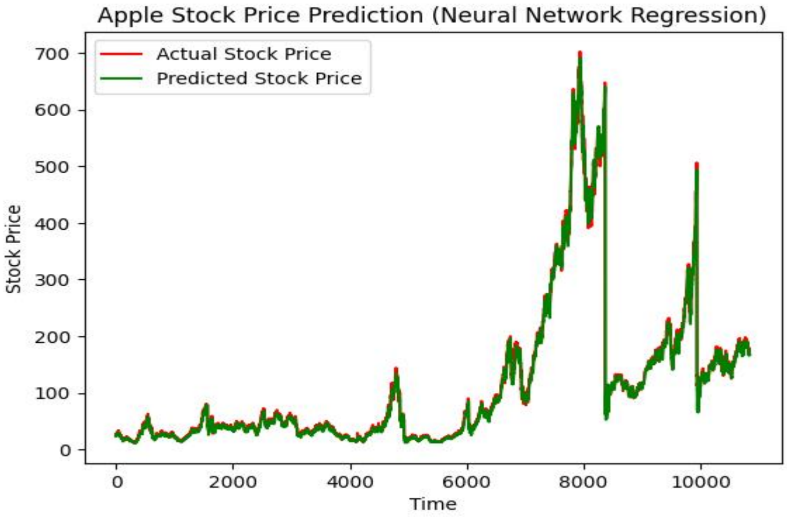


Fig. 3. Plot of Actual Stock Price vs Predicted Stock Price of KNN

Fig. 4. Plot of Actual Stock Price vs Predicted Stock Price of Neural Networks

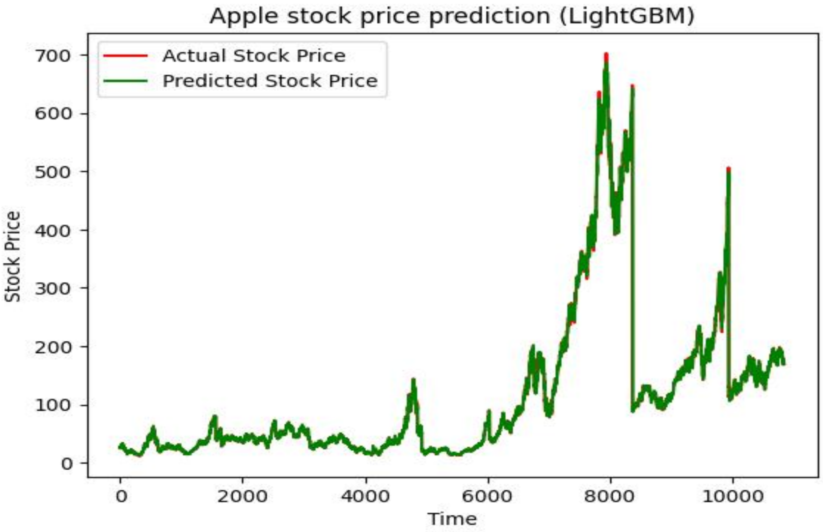


Fig. 5. Plot of Actual Stock Price vs Predicted Stock Price of LightGBM

The obtained R^2 and MSE values of the five algorithms applied to the collected datasets was shown in Table 2. The R^2 value for LSTM is 0.9963214, which means the accuracy level for LSTM is 99.63%. Likewise, the accuracy level for KNN, XGBoost, Neural Network and LightGBM is 99.68%, 99.97%, 99.49% and 99.85%, respectively. From the above results, a XGBoost has the highest R^2 value. The higher the R^2 value, the more accurate the algorithm. The value of MSE for LSTM is 63.401569, for the LightGBM algorithm MSE value is 24.857377, for the KNN algorithm MSE value is 54.263800, for the Neural Network algorithm, the MSE value is 87.946904 and for the XGBoost algorithm MSE value is 5.125808 which is least from all the five algorithms used. The lesser the MSE value the more accurate the algorithm will be. So from the above results, the XGBoost is the best one.

Table 2. Comparison of R^2 and MSE

MODEL	R2	MSE
LSTM	0.9963214	63.401569
KNN	0.996851	54.263800
XGBOOST	0.999702	5.125808
NEURAL NETWORK	0.994900	87.946904
LIGHTGBM	0.998558	24.857377

Conclusion

The following important conclusions from our examination of this study on the use of several machine learning algorithms to anticipate stock market prices: The dataset that was employed in the process of building the model included fundamental financial data, including past stock prices and trade volume.

- To predict stock market prices, six machine learning algorithms were used: XGBoost, LSTM, KNN, Neural Network Regression, and LightGBM.

- Two main measures were used to assess the created algorithms' performance: the Mean Squared Error (MSE) and the coefficient of determination (R2).
- Among the models evaluated, XGBoost showed remarkable accuracy, with an MSE of 5.125808 and an R2 value of 0.999702.
- In addition to XGBoost, LSTM and KNN demonstrated remarkable predictive performance, with R2 values of 0.9963214 and 0.996851, respectively. However, their MSE values were marginally higher.
- With R2 values of 0.994900 and 0.998558, Neural Network Regression and LightGBM produced decent results; nevertheless, their MSE values were greater than XGBoost's.

All things considered, our research shows that XGBoost is better at predicting stock market values than other single algorithms. In order to improve the predictive power of these models and help investors make better-informed investment decisions in the financial markets, more investigation and testing may be necessary.

References

[1] Parmar, R.R., Roy, S., Bhattacharyya, D., Bandyopadhyay, S.K. and Kim, T.H., 2017. Large-scale encryption in the Hadoop environment: Chal

[2] Patel, J.; Shah, S.; Thakkar, P.; Kotecha, K. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Syst. Appl. 2015, 42, 259–268.

[3] Sezer, O.B., Ozbayoglu, A.M. and Dogdu, E., 2017, April. An artificial neural network-based stock trading system using technical analysis and big data framework. In Proceedings of the South East Conference (pp. 223-226). ACM.

[4] Desai, R.; Gandhi, S. Stock Market Prediction Using Data Mining. Int. J. Eng. Dev. Res. 2014, 2, 2780–2784.

[5] Khashei, M.; Hajirahimi, Z. Performance evaluation of series and parallel strategies for financial time series forecasting. Financ. Innov. 2017, 3, 1–24.

[6] Chung, H., & Shin, K. S. (2018). Genetic algorithm-optimized long short-term memory network for stock market prediction. Sustainability, 10(10), 3765.

[7] Kumar, G., Jain, S., & Singh, U. P. (2021). Stock market forecasting using computational intelligence: A survey. Archives of computational methods in engineering, 28, 1069-1101.