Stock Price Prediction Using Machine Learning Algorithms

Aman Mohapatra, Oorja Singh, Prachi Saurabh, Pranay Harchandani, Kalpita Chakraborty
School of Computer Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
2105352@kiit.ac.in, 2105387@kiit.ac.in, 2105389@kiit.ac.in, 21052914@kiit.ac.in, 21052503@kiit.ac.in

Abstract—**In the dynamic world of finance, predicting stock prices accurately paves the way for profitable investments. Machine learning algorithms, with their ability to learn from data and improve over time, have emerged as a powerful tool in predicting stock prices. This study explores the application of various machine learning algorithms for stock price prediction. The algorithms analyzed include Linear Regression, Polynomial Regression, Decision Trees, K Nearest Neighbors, Recurrent Long Short Term Memory(RLSTM), Long Short Term Memory(LSTM). These algorithms are trained on historical stock data, learning patterns and making predictions about future prices. The performance of each algorithm is evaluated based on its prediction accuracy. The results indicate that machine learning algorithms can provide significant insights for stock price prediction, aiding investors in making informed decisions. However, it's important to remember that while these predictions can be highly accurate, they're not foolproof. The stock market is influenced by a myriad of factors, and even the most sophisticated algorithms can't account for all of them. Therefore, these predictions should be used as one of many tools in an investor's toolbox.**

Keywords—stock price prediction, machine learning algorithms, Linear Regression, Polynomial Regression, K Nearest Neighbors, Decision Trees, RLSTM, LSTM, Accuracy, historical stock data.

I. INTRODUCTION

In the intricate world of financial markets, stock analysis stands as a beacon, guiding investors through the complexities of trading instruments, investment sectors and markets as a whole . It is a process that seeks to predict future market activities by studying past and current data, thereby enabling investors to make informed buying and selling decisions . The stock market , a key pillar of global economy, provides businesses with the necessary funding for growth and expansion. The concept of stock price analysis hinges on the assumption that market data can be used to determine a stock's intrinsic value . However, accurately estimating stock prices is a challenging task due to volatile nature of the markets, which are influenced by a multitude of factors including political events , economic indicators and company performance.

In high stakes arena of stock market investments, the ability to accurately predict the performance of individual stocks is of paramount importance to investors. This is particularly true for companies like Tesla inc., a pioneer in electric vehicle and clean energy industry. The application of analytical techniques is crucial given Tesla's innovative endeavors and the inherent volatility of the stock market.

Machine Learning techniques can use historical price and volume of data to identify patterns and trends that may indicate future price movements. By interpreting these signals , investors can identify potential entry and exit points in Tesla's stock . This approach humanizes the complex process of stock analysis , making it more accessible and understandable for the investors.

II. RELATED WORK

Srinath Ravikumar and Prasad Saraf (2020) have hypothesized and drawn a comparison among various different algorithms such as Support Vector Machines, Random Forest Classification among others on the basis of their accuracy and execution time. They have worked on their paper by an implementation through Spyder. Lastly, they have made a Confusion Matrix to further assert the performance of the methods they have used in their paper.

Saleh Alhazbi, Ahmed Ben Said, Alanoud Al-Maadid (2020) have used convolutional neural network (CNN) to predict stocks direction in Qatar Stock Exchange (QE) as a case of emerging rates. They have included historic data along with S&P index, Nikkei index and oil prices as features which increased the accuracy of their model by 10%.

Nagaraj Naik, Biju R. Mohan (2021) have proposed a stock crisis prediction model based on Hybrid Feature Selection (HFS) technique, Naïve Bayes, Relative Strength Index (RSI) method to find a bubble in the stock price, moving average statistics to identify crisis point in stock prices, Extreme Gradient Boosting (XGBoost) and Deep Neural Network (DNN) regression method. They evaluated the performances based on MSE, MAE and RMSE.

Louis Owen, Finny Oktariani (2020) have investigated the potential of exploiting sentiment score extracted from microblog text data along with historical stock data to improve stock market prediction performance using ensemble-based model that use LSTM, MLP and CNN. They proposed a Stock Ensemble-based Neural Network (SENN) and evaluated their models based on Adjusted MAPE metric.

Sona Susan Jacob, Sankha Patra, Kapinesh G, Thanikaiselvan V (2020) have developed a dashboard for monitoring stock prices in real time and an LSTM model to forecast stock prices and have compared its performance with some machine learning algorithms.

Shweta Agarwal, Utkarsh Goel, Shailendra Kumar (2021) have attempted to investigate the impact of information diffused through the news media on the stock markets of an emerging economy, with evidences from India.

III. METHODOLOGY

The functioning of the system is mentioned below:

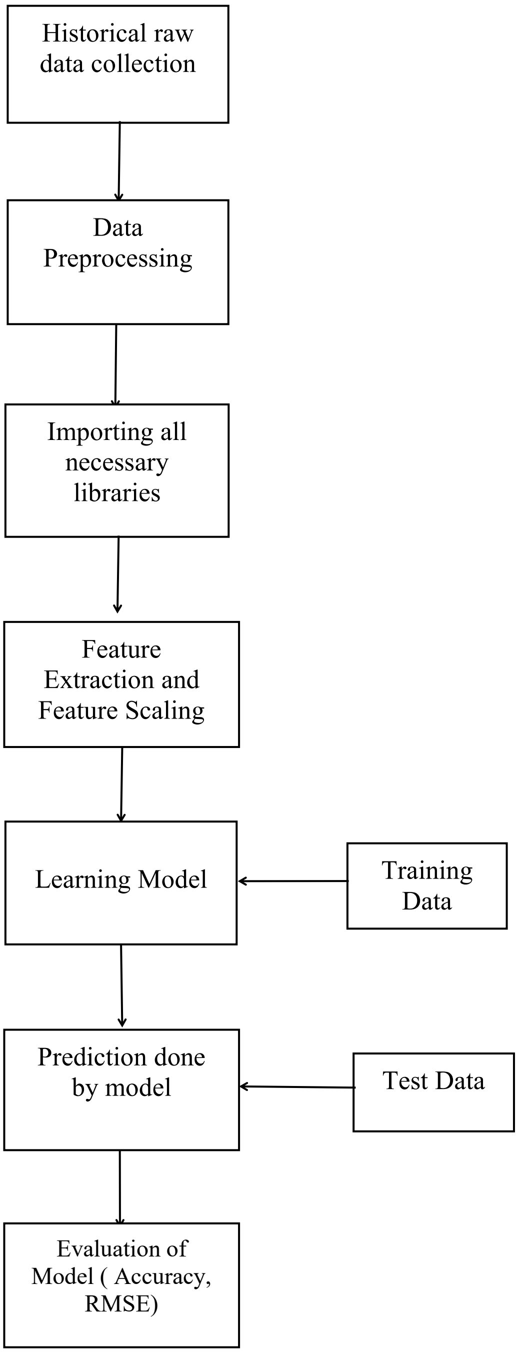

Fig. 1. Flowchart of application

1. The raw data used is from Tesla
2. The attributes used for feature extraction are 'date' , 'volume', 'closing price' of a stock.
3. The dataset is split into training and test data set..
4. The traning dataset is used for model training and test dataset is used for prediction.
5. The values of the test data are predicted and the results are evaluated . The result is given on the basis of Accuracy and Training Time.

IV. ALGORITHMIC APPROACH

To initiate the development of a machine learning model, we require a dataset first. Normally, open-source data on the internet might have some issues like missing values, duplicates or no structure. The reason why this data needs to be pre-processed before being used for training a model is to ensure its quality and appropriateness.
Financial datasets such as historical stock price data usually have several major attributes which include:

Date: This refers to the date corresponding to a specific stock price.

Opening Price: This implies the initial price of the stock on a given day.

High Price: It represents the highest value reached by the stock during the trading day.

Low Price: It shows the lowest value reached by the stock during the trading day.

Closing Price: The final price of a share when trading ends,taken into account as most important attribute for modeling purposes.

Volume: Total number of shares traded in one day.

Out of these attributes,the closing price is often regarded as being most significant for models' use.What we do is that after investigating past closing prices ,we can utilize different regression models in machine learning so that they help us predict future stock prices.In regression,a curve is plotted on a graph based on input data where dates are represented along x-axis while closing prices are represented along y-axis.

The models used in this dataset are:
1. Simple Linear Regression
2. Decision Tree Classification
3. KNN Classifier
4. RLSTM(Recurrent Long Short-Term Memory)
5. LSTM (Long Short-Term Memory)

The models' performance are evaluated based on two measures:

1. Accuracy: Accuracy is one of the most widely used performance measures in Machine Learning. This method is extremely straightforward and quantifies the number of correct predictions made out of the total number of predictions made. In other words, accuracy is a metric that measures how often a model's predictions are correct.
It is given mathematically as,
Accuracy = (TP+TN)/(TP+TN+FP+FN)
Where,
TP stands for **true positive** and is the number of times the model correctly predicts the outcome to be positive.

TN stands for **true negative** and is the number of times the model correctly predicts the outcome to be negative.
FP stands for **false positive** and is the number of times the model incorrectly predicts the outcome to be positive.
FN stands for **false negative** and is the number of times the model incorrectly predicts the outcome to be negative.

The higher the accuracy , the better is the model's performance.

2. Training Time: Training time is the time taken by the model to learn the optimum weights and bias for the given data based on the training examples supplied.

The lower the training time , the better is the model's performance.

V. Implementation and Result Analysis

1. Simple Linear Regression

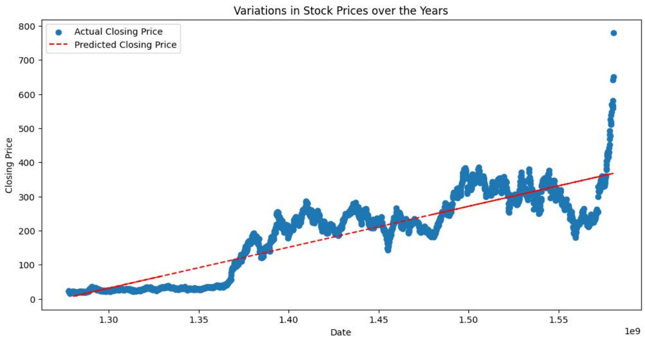

Fig. 2. Simple Linear Regression

2. Decision Tree Classifier

Fig. 3. Decision Tree Classifier

3. KNN Classifier

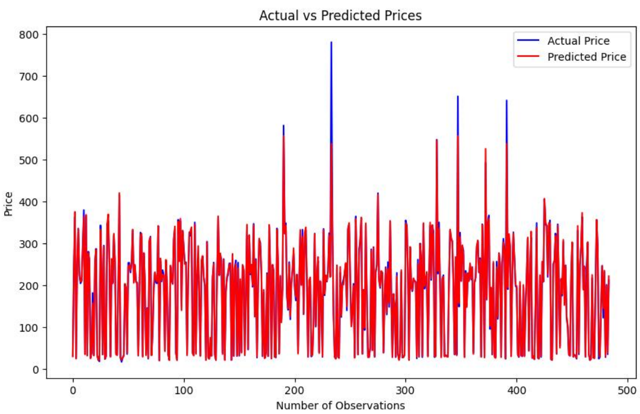

Fig. 4. KNN Classifier

4. RLSTM

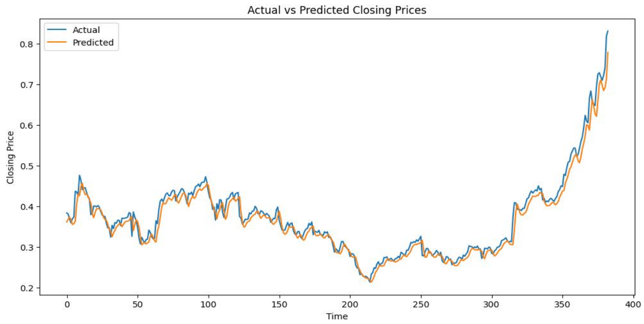

Fig. 5. RLSTM

5. LSTM

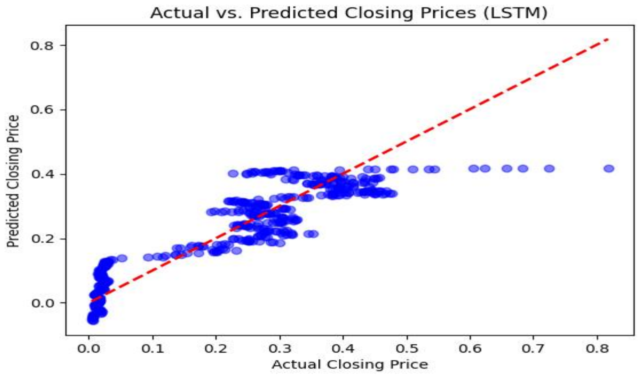

Fig. 6. LSTM

TABLE I. Result Analysis of Models

Model	***Accuracy*** *(within 5% threshold)*	***Training Time***
Linear Regression	0.10	0.00261735916137695
Decision Tree	0.7128099173553719	0.07370185852050781
KNN Classifier	0.8429752066115702	0.00825238227844238
RLSTM	135.78067885117494	220.4039716720581
LSTM	95.80165289256199	15.52771544456482

VI. Conclusion

The dataset obtained from Tesla Inc was efficiently preprocessed . The Training time and accuracies of the various machine learning models were compared and analyszed .
The RLSTM model gave the maximum accuracy .

Acknowledgement

We would like to thank Dr. Abhaya Kumar for his guidance and support in making this paper. This paper was made after taking inspiration from a lot of papers referenced below. We would also like to thank our institution, the School of Computer Science, Kalinga Institute of Industrial Technology, for organizing this research.

REFERENCES

[1] Srinath Ravikumar, Prasad Saraf, “Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms” 2020 International Conference for Emerging Technology (INCET), June 2020

[2] Saleh Alhazbi, Ahmed Ben Said, Alanoud Al=Maadid, “Using Deep Learning to Predict Stock Movements Direction in Emerging Markets: The Case of Qatar Stock Exchange” 2020 International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), February 2020

[3] Nagaraj Naik, Biju R. Mohan, “Novel Stock Crisis Prediction Technique—A Study on Indian Stock Market” IEEE Access, vol. 9, pp. 86230 – 86242, June 2021

[4] Louis Owen, Finny Oktariani, “SENN: Stock Ensemble-based Neural Network for Stock Market Prediction using Historical Stock Data and Sentiment Analysis”, 2020 International Conference on Data Science and Its Applications (ICoDSA), August 2020

[5] Sona Susan Jacob, Sankha Patra, Kapinesh G, Thanikaiselvan V, “Monitoring of Stocks using LSTM Model and Prediction of Stock Prices” 2022 Internation Conference on Edge Computing and Applications (ICECAA), October 2022

[6] Shweta Agarwal, Utkarsh Goel, Shailendra Kumar, “News media sentiments and Stock markets: the Indian perspective” 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), December 2020