**Stock Price Prediction Using Machine Learning Algorithms**

Aman Mohapatra1, Oorja Singh1, Prachi Saurabh1, Kalpita Chakraborty1, Pranay Harchandani1

1School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar,

Odisha

{amanmohapatra50, oorjasingh2105, prachismssaurabh, pranayharchandani5@gmail.com}@gmail.com

***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**

**1 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.

**2 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 networks (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 scores extracted from microblog text data along with historical stock data to improve stock market prediction performance using ensemble-based models 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 evidence from India.

**3 Dataset**

We are using a dataset of Tesla to predict the stock prices. The original dataset contains Date, Open, High, Low ,Close, Adj Close, Volume as the attributes but here ‘date’ , ‘volume’, ‘closing price’ of the stock is used for feature extraction.

**4 Methodologies**

The functioning of the system is mentioned below:

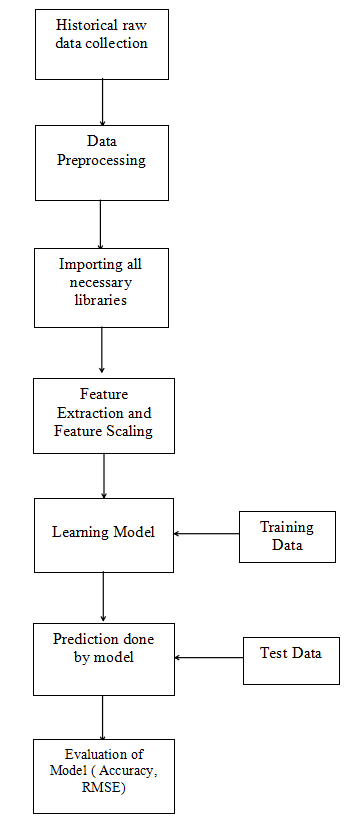


Fig 1: Flowchart of application

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 dataset.
4. The training dataset is used for model training and the 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

The model used in the dataset are:

1. Simple Linear Regression
2. Decision Tree Classification
3. KNN(K-Nearest Neighbour) Classifier
4. RLST(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.

**5. Results and Discussions**

1. Simple Linear Regression

**References**

1. Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195--197 (1981)

2. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148--1158. Springer, Heidelberg (2006)

3. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)

4. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181--184. IEEE Press, New York (2001)

5. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration. Technical report, Global Grid Forum (2002)

6. National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov

**Appendix: Springer-Author Discount**

LNCS authors are entitled to a 33.3% discount off all Springer publications. Before placing an order, they should send an email to [orders-HD-individuals@springer.com](mailto:orders-HD-individuals@springer.com), giving full details of their Springer publication, to obtain a so-called token. This token is a number, which must be entered when placing an order via the Internet, in order to obtain the discount.

**Checklist of Items to be Sent to Volume Editors**

1. A final Word or RTF file
2. A final PDF file
3. A copyright form, signed by one author on behalf of all of the authors of the paper
4. A readme giving the name and email address of the corresponding author