**LOSS PREVENTION SYSTEM IN STOCK MARKET TRADING USING MACHINE LEARNING**

Sambangi Mounika, Ambitama Poddar, Dr. Chandra Mouliswaran Subramanian

*School of Information Technology &Engineering, Vellore Institute of Technology*

*VIT, Vellore Campus, Tiruvalam Rd, Vellore, Tamil Nadu,632014*

sambangi.mounika2022@vitstudent.ac.in

ambitama.poddar2022@vitstudent.ac.in

chandramouliswaran@vit.ac.in

***Abstract-*** Stock market prediction helps determine future stocks and maximize profits. Our purpose is to have smooth functioning of trading. Every company has its ups and downs daily, so it is necessary that investors must be able to predict market movements accurately. So, the Machine Learning algorithms took place to reduce the risk involved in trading activity and make it less complex. The paper mainly focuses on the regression and long short-term memory (LSTM) models. We get to know the total number of shares, including their prices and volume breakouts. We use the Auto-Regressive Integrated Moving Average (ARIMA) model for utilizing historical data on price and volume. When we work with non-stationary data, the ARIMA model converts into stationary data. It is one of the most widely used models for predicting linear time series data. The Factors considered are open, close, low, and high prices for a given period. The models are evaluated using the indicators RMSE. We will be comparing the accuracy of the respective models and finding the best model to predict stock prices. The main aim is to obtain greater accuracy by successful prediction and maximize profits.

***Keywords-*** Stock market; Old stock details; Livestock details;

Volume breakouts; Regression

1. **INTRODUCTION**

A successful forecast of stock prices will be expected by investors, sellers, and brokers which will result in enormous gains. Normally the predictions are often completely confused, so they should be made by examining the history of the stock market properly. Then it is necessary to take a machine learning technique to do the implementation. The predicted value should give similar results as of original stock values then we can improve the accuracy by choosing an even more suitable machine learning algorithm and getting the best fit to the data. So, a lot of studies have to be drawn to get effective and precise measures regarding stock market prediction.

A key role is played in collecting the datasets. A small modification in the data might lead to misleading results. So necessary steps are taken to make it as precise as possible. We collected the dataset from Yahoo Finance. This dataset includes the following five parameters: open, close, low, high, and volume stock prices. In a particular time period, the total number of shares transferred between one owner to another is the volume. Next, the model is used for testing the data.

For this hypothesis, LSTM, ARIMA, and regression models are used individually. The regression includes reducing error, LSTM helps with long-term memory retention of the data and findings and ARIMA for predicting the time series data. The graphs showing price fluctuations with dates and between actual and expected prices are shown here as a final step.

Today, stock price predictions are being made using the benefits of technology methods. This is due to the fact that the amount of data needed for forecasting grows rapidly yet does not follow a linear trend. To manage this new data dimension, artificial intelligence-supported models that can adapt to complicated interactions are required. The support vector

machine, random forest approach, and regression analysis (attribution) are the techniques that are most frequently employed in this subject. The estimation of stock prices depends on many parameters which makes it difficult to handle. However, choosing the correct parameters and integrating them into the analysis phase can eliminate this problem. Therefore, we used two different indicators in our study. These are the "Aroon Oscillator" and the Stochastic Momentum Index.

1. **RELATED WORK**

(Fischer & Krauss, 2018) [3] collected the data set from S&P from Thomson Reuter between December 1992 to October 2015. Usage of LSTM networks and models like the random forest, DNN, and a logistic regression classifier (LOG). They converted the lists into the binary matrix and then divided the data into training and testing sets. They used only selected models and found the best one to predict the stocks.

The dataset from the daily Istanbul Stock Exchange from January 2, 1997, to December 31, 2007, was collected. (Kara, Boyacioglu, & Baykan, 2011) [5] used ANN and SVM techniques. The parameters were set to start the experiment. They tested 900 parameters for the ANN model. The performance is checked during training and testing. Since ANN is used, it had its own limitations in predicting the stocks.

The dataset was collected from Nikkei 225 index from January 2007 to December 2013. (Qiu & Song, 2016) [8] used the techniques of ANN and Genetic Algorithms. The GA- ANN hybrid model is used to compare predictions with actual data. 78.6% was used as a training set and 21.4% are considered as testing data. To overcome the nonlinear optimization problems that occurred using ANN, we use a global search technique such as GA. We use GA to optimize the weights and biases of the ANN.

They collected the data set from the Korean stock price index a sample from January 1989 to December 1998. (. Kim & Shin, 2007) [1] used artificial neural networks with genetic algorithms to detect the patterns. They applied GAs for optimization of the number of time delays for ATNN and TDNN models. The results showed that the accuracy of this proposed model is higher than that of ATNN, TDNN, and RNN. One of the methods of ANN is BPN which learns only static patterns that are not dependent on time. To overcome this limitation, they used recurrent links and time-delayed links.

The dataset from the NASDAQ index of the Taiwan Economic Journal Database from 2008 and 29 technical indices was used. (LEE, 2009) [6] used a model based on a support vector machine with a hybrid feature selection. Supported sequential forward search used in the feature selection process. The advantage of this is to provide a detailed procedure to adjust parameters where we can evaluate the performance at different parameter values. They only compared back-propagation neural networks and no other machine learning algorithms.

The dataset from the Korea Stock Price Index from January 2000 to December 2016 was used. (Chung, 2018) [2] used a model based on RNN using LSTM, which is a very important method in deep learning. They used GA and LSTM models and solved using various parameters. 80% of the data was used to train and 20% is considered as testing data. The limitations can be seen in maintaining linearity and normality. To overcome this, GA algorithms should be adapted.

(ARAVIND GANESAN, 2021) [11] Predicted the stock prices of ICICI Bank and Reliance Industries using the ARIMA model and the historical stock data of ICICI Bank and Reliance Industries have been collected from Yahoo Finance. They used the Arima model for stock price prediction. They used various packages in python along with the ARIMA model to project the future values of time series. The data is split into two parts, 70% to train and 30% to test the model. After implementing the ARIMA model, with optimal values, they depicted a plot that contains the Actual price and predicted price of the stock. The drawback of this analysis is that the ARIMA model holds higher accuracy for short-term predictions.

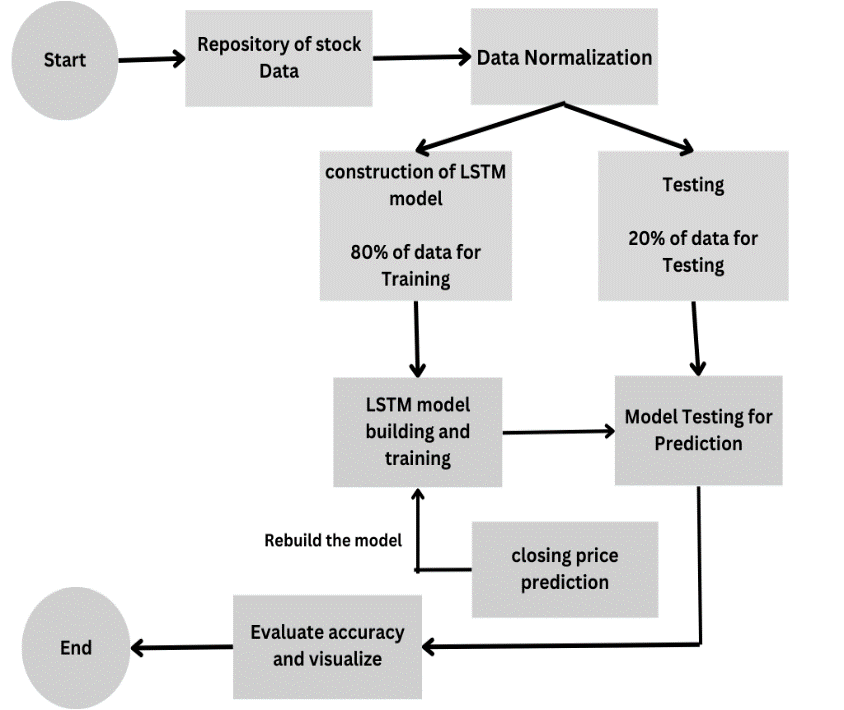
 (Mohanty., 2021) [7] took three years of data collected on a monthly basis from 18th April to Feb 21. ARIMA model was used to help investors to predict stock values. They used python software to apply the ARIMA model and successfully predict the prices of three sectors. They converted the non-stationary statistic to stationary statistic by changing the closing price in the log and the predicted prices are shown in the form of a graph. The drawback of this analysis is that the ARIMA model holds higher accuracy for short-term predictions.

Figure 1: Architecture of stock price prediction

The dataset collected from NASDAQ daily stock exchange from October 7, 2008,

to June 26, 2009, is used. The models analysed by (Guresen, Kayakutlu, & Daim, 2011) [4] are MLP, DAN, and use GARCH to extract new input variables. By using data from the NASDAQ Stock Exchange index, they implemented these models and calculated mean square error and mean absolute deviation for the data. It didn’t show consistent results in the regression model by using DAN.

The dataset from the monthly Istanbul stock exchange for the period from 2009 January to 2021 March. (Tamerlan Mashadihasanli) [10] used the ANN model to forecast stock market prices and another model is ARIMA for forecasting the time series. They found it to be more efficient. They took 147 observations to evaluate and forecast the model. The data is divided to train to estimate the model, and the other part is to test used for forecasting. Their result shows positive skewness has a high positive kurtosis The number of indicators in the obtained dataset is limited and it should be increased and done with high accuracy.

3. **Modelling Approach**

3.1 **Data Collection and Preparation**

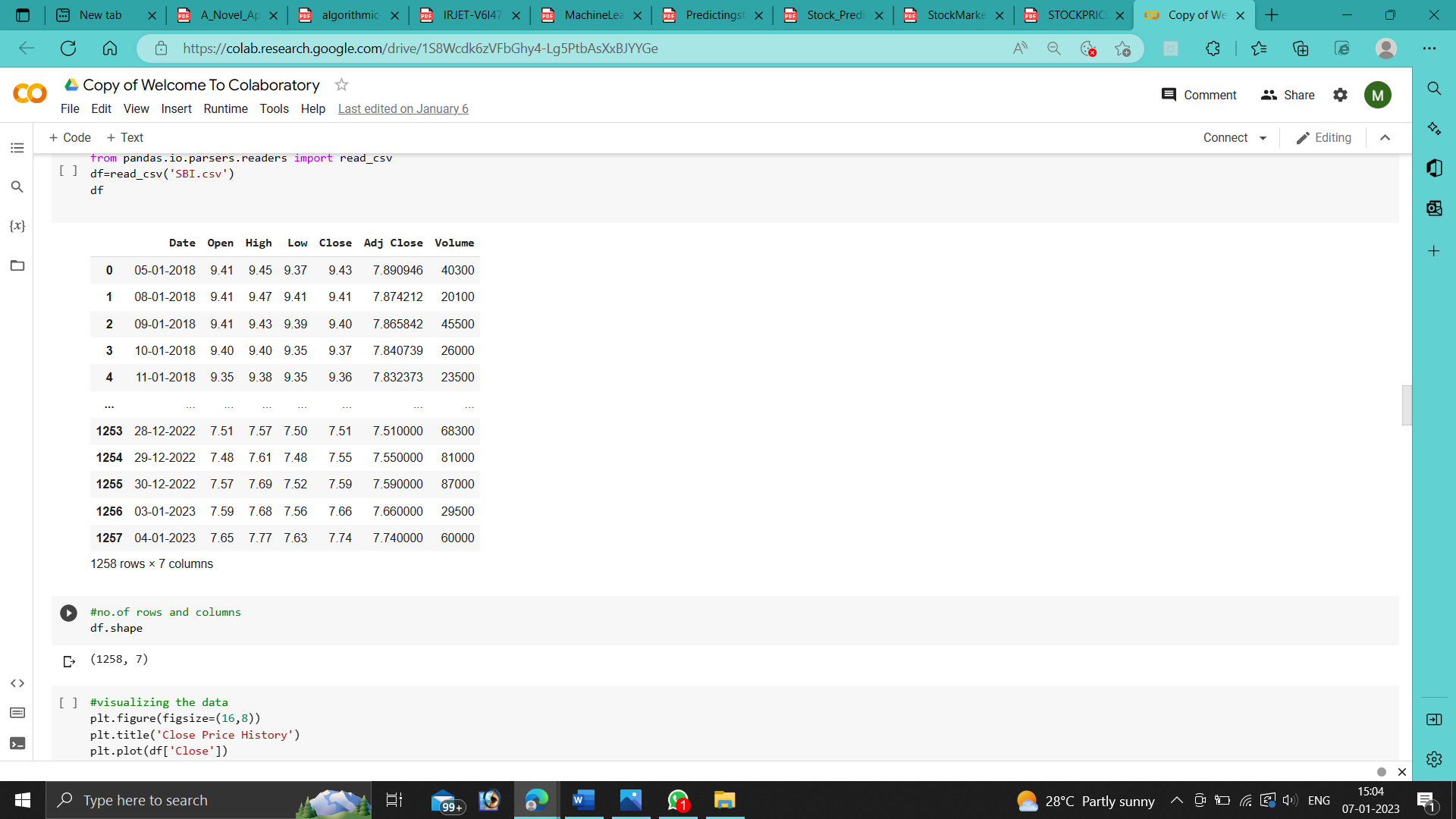
The dataset collected from yahoo finance includes data from the banking sector for around 10 years from 1/1/2012 to 31/12/2022. From below table 1, I selected the close prices column to train and test the model.

Table 1: Collection of Data

The original data should be divided into training and testing data sets and then the results are measured with actual values. I separated the data in an 80:20 ratio, where 80% will be used for training the data set and 20% is used for testing the data set. To simplify the algorithm’s identification of patterns in the data, the data has been scaled to standardize it within a range of 0 to 1.

I built the training dataset with a timestamp of 60 after scaling the data. For each repetition of input and output, the model processes data in batches of a certain size. As a result, it will utilize the first 60 data as input and the 61st as output in the first iteration. The second iteration uses the second through 61st as input, the second through 62nd as output, and so on. I create the LSTM model in the next phase.

Epoch is the total number of iterations the model makes through the entire training set. I'm using epoch number 10 in this case, which means the model will run through each dataset 10 times.

* 1. **RNN**

LSTM is one of the methods of RNN which uses past records and helps to predict future prices. It is more effective in storing the previous memory mainly in the case of huge datasets.

The major drawback of RNN is that it encounters two problems during training in the case of large data sets.

* **Vanishing Gradient**

Vanishing gradients occur when the gradient of weights with respect to the loss function is very small and even decreases the learning rate i.e. it can even disappear thus slowing the training of the model. So it’ll be hard to learn the parameters of previous layers. We can use backpropagation and update the weights to reduce the loss of the model.

* **Exploding Gradient**

Exploding gradients occur when gradients of weights with respect to loss function are very large, which causes the model to become unstable and lead to poor performance, it can even explode. It can be caused by the activation function.

1. **Proposed System**

The starting stage is to include historical data in our implementation. The second step is removing unnecessary columns and taking only the required features to analyse the data. Then we should train the data sets and then test the results with the remaining data which helps in predicting the prices and then at last we’ll view the results. The suggested design for the implementation of the system is shown in the below Figure.

Step 1: Collection of data from any source

Step 2: Data should be preprocessed

Step 3: Scaling of values in data between 0 and 1

Step 4: Divide data into the training set and test set

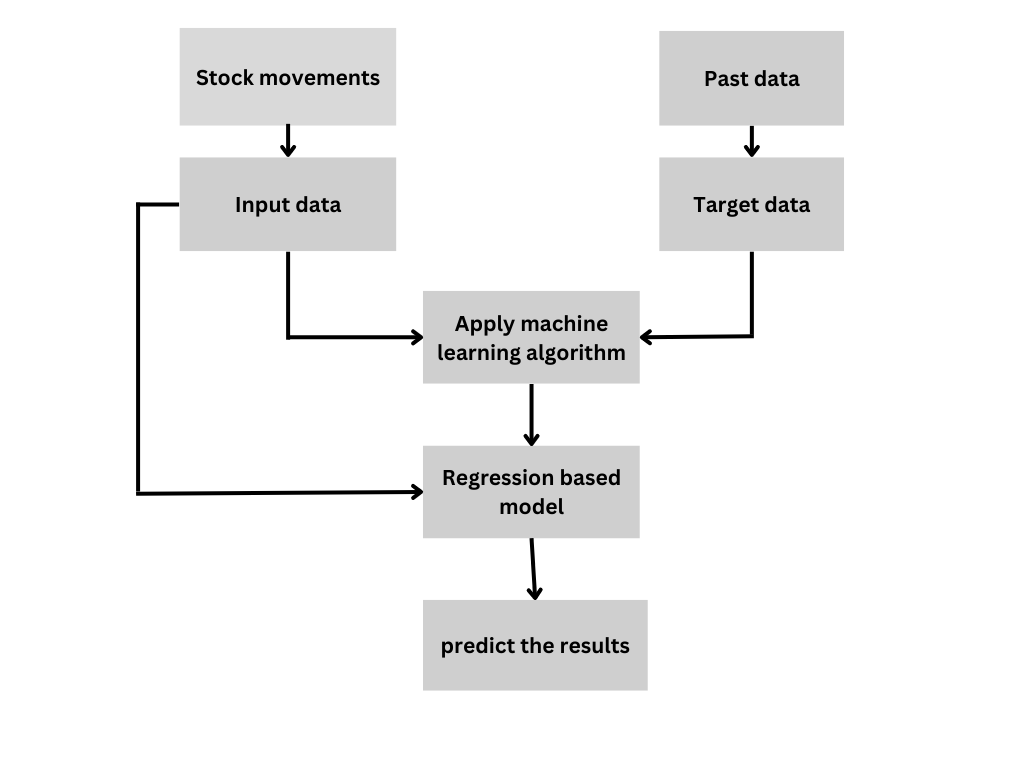
Step 5: Build the LSTM model

Step 6: Training the model

Step 7: Make predictions of stock prices

Step 8: Compare true versus predicted adjacent close values by plotting the graph

Step 9: Evaluate the accuracy and visualize the result

Figure 2: Proposed System for stock price prediction

**4.1 Long Short-Term Memory (LSTM)**

There are issues like vanishing gradient and exploding gradient in RNN, where we are using one of the methods of RNN i.e., long short-term memory to address this issue. This type of RNN is unique in its kind that can pick up on long-term dependence. This is made to prevent difficulties with long-term dependence and help people retain information of past data for extended periods. The LSTM employs supervised learning and it helps in processing the datasets of the entire sequence.

**4.1.1 Construction of the LSTM model:**

We will need to import the Keras library and all other packages in order to construct the LSTM model. Keras is a high-level TensorFlow API for creating and refining models. Here, we'll import a few libraries.

We also use some models like:

Sequential - It is essentially a linear stack of layers through which the list can be passed to produce a sequential layer that helps us access each step at each hidden layer.

Dense is a fully connected layer type where each neuron of a layer will be connected to neurons of the other layers. The number of units mentioned will have an effect on the output shape of the dense layer.

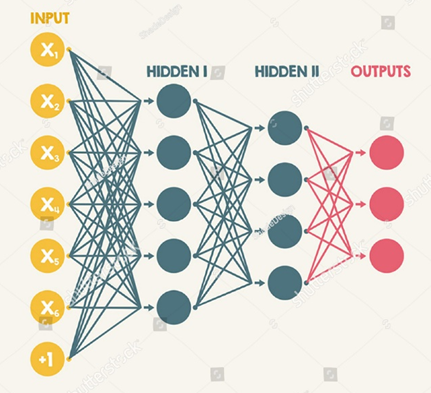


Figure 3: LSTM

For the building of the LSTM model, we are using sequential layers, dense layers, and many others. After that, we need to compile the model using an optimizer. Here we are using Adam optimizer. We need to calculate the root mean square error.

Then we need to train the model by giving the batch size and the number of epochs. Then create a testing data set and convert data into NumPy arrays and then reshape the data and get predicted price values and root mean square error. The amount of error percentage shows the effectiveness and accuracy of the model.

* 1. **ARIMA Model**

ARIMA model is used to analyse and predict time series of data. ARIMA refers to Auto-Regressive Integrated Moving Average. By using this we can convert non-stationary data into stationary.

In the ARIMA model, we are using three ordered parameters (p, d, q) and these parameters are filled with integer values. These parameters are the coordinates of an ARIMA model. The tuning parameters can be determined from the graphs of autocorrelation and partial autocorrelation.

• The term p refers to the number of autoregressive terms or the number of observations from past time values utilized to predict future values.

d-The time series must have d differences in order for it to remain stationary that includes constant mean, variance, and autocorrelation

• q or the lagged values of the error term, is the moving average of the prior forecast mistakes in our model. As an illustration, if the value of q is 1, the error term in the model has one lagged value. By using the parameters in the best fit of the Arima model and predicting future stocks.

**4.3** **Regression model**

In every daily field, linear regression plays an important role in predicting and forecasting certain things. It determines the relationship between independent and dependent variables.

The linear regression model can be used to predict the prices of stock using some machine learning techniques. This model will help in providing the best fit of the line. T contains an independent variable, constant or y-intercept, slope, and error term or residuals. Now we’ll use a linear regression model to fit both input and output data and then predict. Finally, plot the graph between closing prices and the regression line. Thus, compare the results.

When finding trends like seasonality or weather patterns, autocorrelation analysis is helpful. However, it is troublesome when attempting to extrapolate results for price prediction. The lesson learned from this is that we need to find another independent variable because our date values aren't appropriate and use the adjusted close value as the independent variable.

1. **Results and Discussions**

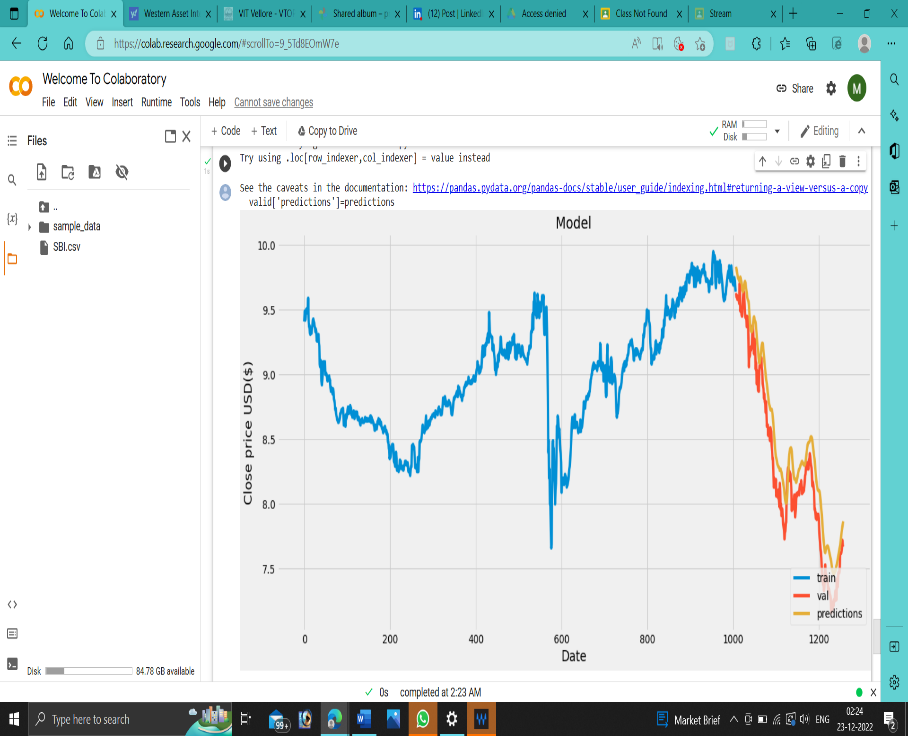


Figure 4: Plot between Date and Price using LSTM

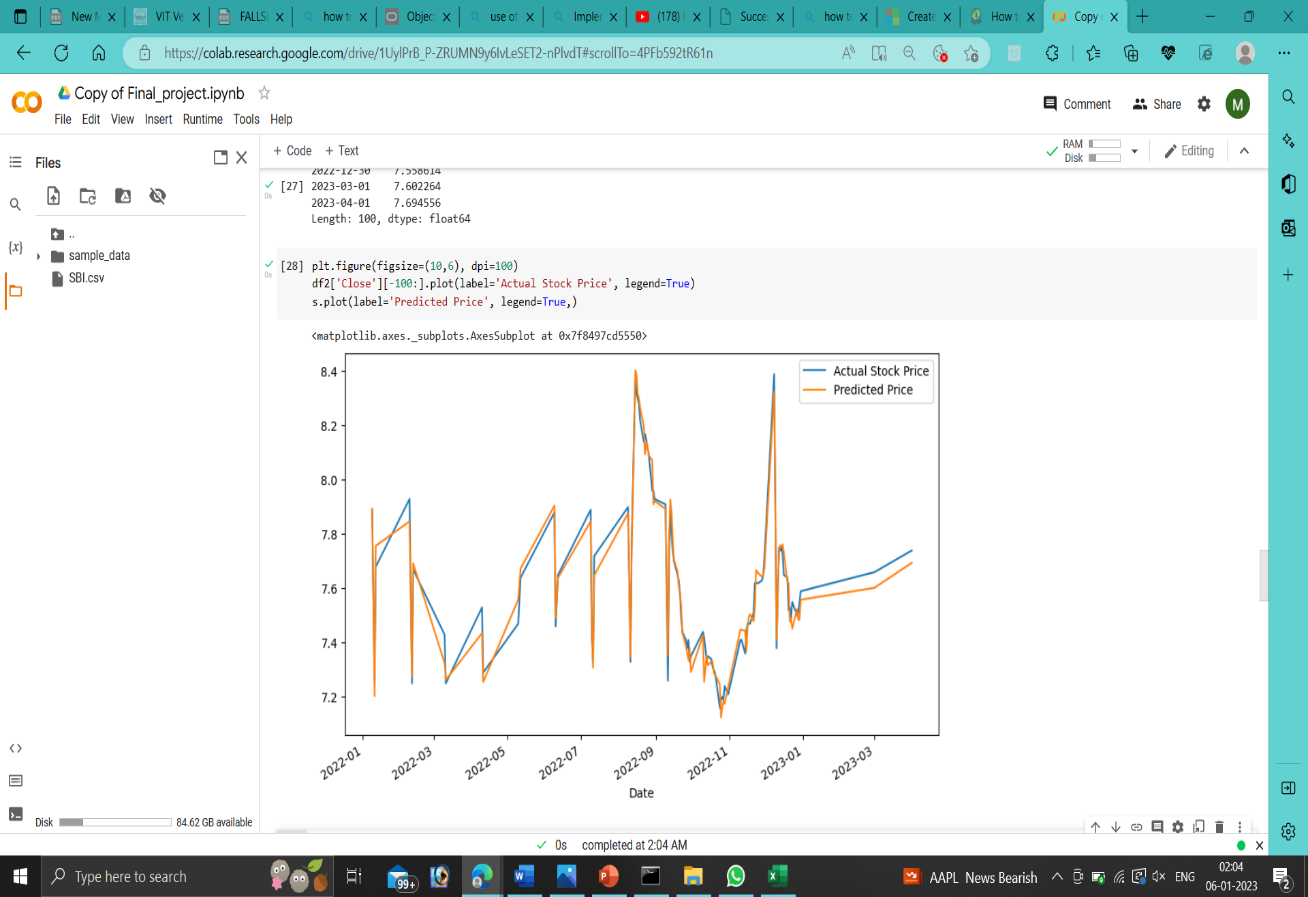
The prediction of stock prices is done using LSTM, where we considered some of the data to train which is in the blue line. Then the predicted values and original values are compared to calculate the accuracy of stock prices obtained as shown in figure 4.

Figure 5: Plot between Actual stock price and predicted Price using Arima

The prediction of stock prices is done using ARIMA, where we considered some of the data to train i.e. Actual stock price which is in the blue line. Then the predicted values and original values are compared to calculate the accuracy of stock prices as shown in figure 5.

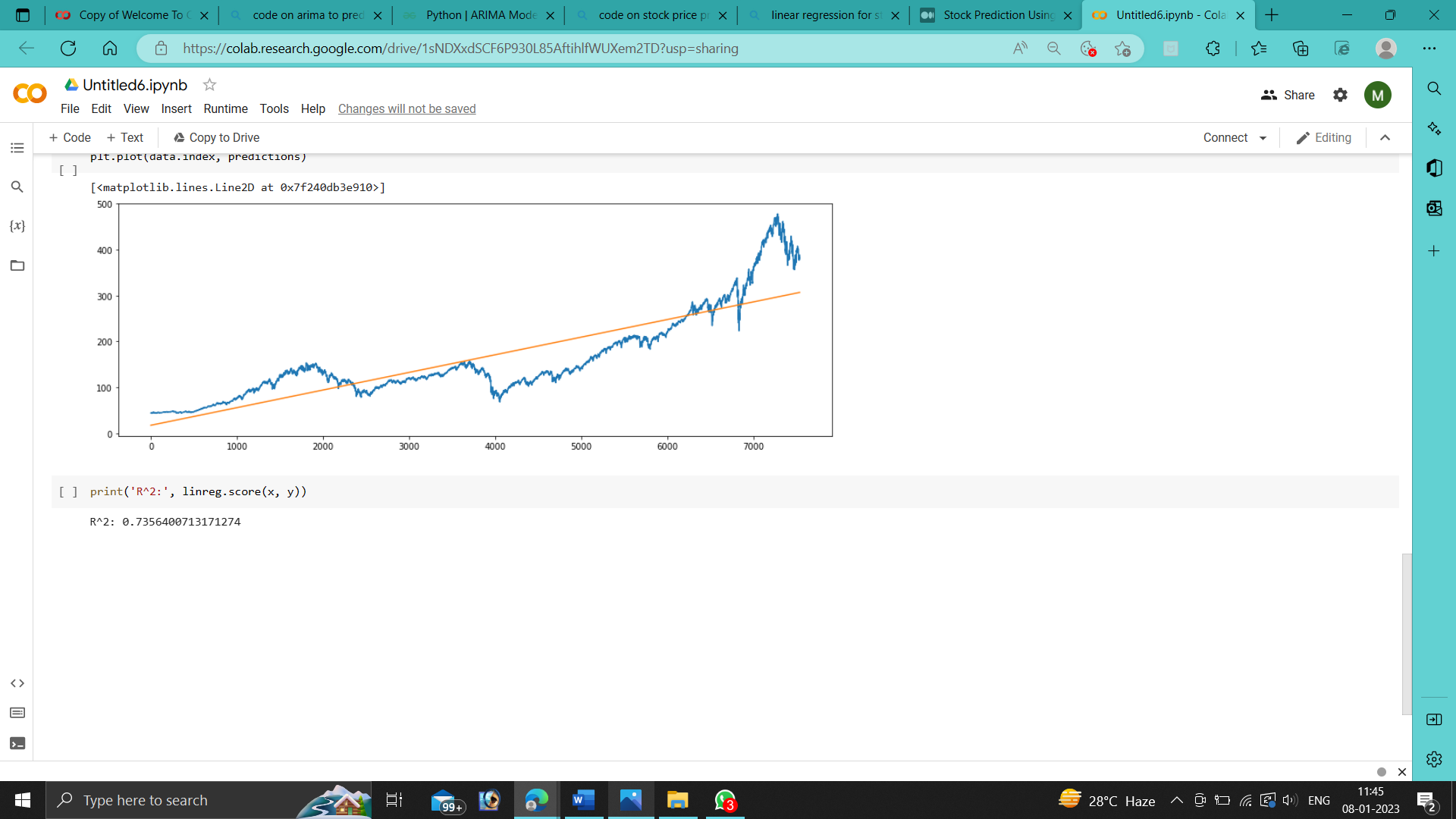


Figure 6: Plot Price and Date using Regression

From the above graphs, the blue line shows the Actual stock price and the yellow line shows the predicted price which is plotted over data having 10 epochs. In all three systems, the predicted price is attained with greater accuracy as shown in figure 6.

|  |  |  |
| --- | --- | --- |
| Linear Regression | ARIMA | LSTM |
| 0.735 | 2.123 | 0.099 |

Table 2: RMSE Comparison with other models

RMSE is used to calculate errors in the respective model and it is calculated by the formula square root of the sum of the square of deviations divided by n. It helps in fitting the best parameters of given data.

1. **Conclusion and Future Work**

In the conclusion, we can say that among the techniques, LSTM, ARIMA, and Linear regression are used and compared the results. The results obtained are satisfactory and hence it shows that it is possible to predict stocks with more accuracy using techniques of machine learning. Here in the above case, the root mean square error of LSTM is relatively low when compared to other models.

In the research, we did use some machine learning techniques. By considering the results of the graphs and calculating the error, we will be getting accuracy. So, it is confirmed that stock prices can be predicted. In future work, we will use other models to get the best parameters to fit the data and improve the accuracy of the prediction.

**Acknowledgments**

We are thankful to all our Vellore Institute of Technology (Vellore) Officials for supporting us and helping us in gaining knowledge for this research work.

**References**

1. Kim, H., & Shin, K. (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. 569–576.

DOI:[10.1016/j.asoc.2006.03.004](http://dx.doi.org/10.1016/j.asoc.2006.03.004)

1. Chung, H. a.-s. (2018). “Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock MarketPrediction.” <https://www.semanticscholar.org/author/Hyejung-Chung/114690062>
2. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. Eur. J. Oper. Res. 654-669.

DOI: [10.1016/j.ejor.2017.11.054](https://doi.org/10.1016/j.ejor.2017.11.054)

1. Guresen, E., Kayakutlu, G., & Daim, T. (2011). Guresen, E.; Kayakutlu, G.; Daim, T.U. Using artificial neural network models in stock market index prediction. Expert Syst. Appl. 10389–10397.

https://doi.org/10.1016/j.eswa.2011.02.068

1. Kara, Y., Boyacioglu, M., & Baykan, Ö. (2011). Predicting the direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Syst. Appl. . 5311-5319.

https://doi.org/10.1016/j.eswa.2010.10.027

1. LEE, M. (2009). Using a support vector machine with a hybrid feature selection method for the stock trend prediction. Expert Syst. Appl. . 10896-10904.

https://www.semanticscholar.org/author/Ming-Chi-Lee/2108722277

1. Mohanty., S. K. (2021). Stock price prediction using arima model. International Journal of Marketing & Human Resource Research.

https://doi.org/10.47747/IJMHRR.V2I2.23

1. Qiu, M., & Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model.

DOI: [10.1371/journal.pone.0155133](https://doi.org/10.1371/journal.pone.0155133)

1. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. . 1929-1958.
2. Tamerlan Mashadihasanli. (n.d.).(2011) Stock Market Price Forecasting Using the Arima Model. Istanbul Stock Exchange. Expert Syst. Appl. , 38, 5311–5319.

https://doi.org/10.26650/JEPR1056771

1. ARAVIND GANESAN, A. K. (2021). Stock Price Prediction using ARIMA Model.
2. Lili Chen, L. S.-M.-E.-T. (2021) ; Stock Trading System Based on Machine Learning and Kelly Criterion in the Internet of Things

<https://doi.org/10.1155/2021/7632052>

1. Strader, Troy J.; Rozycki, John J.; ROOT, THOMAS H.; and Huang, Yu-Hsiang (John) (2020) "Machine Learning Stock Market Prediction Studies: Review and Research Directions," Journal of International Technology and Information Management: Vol. 28 : Iss. 4, Article 3.
2. Shreya pawaskar,(2022). Stock price prediction using machine learning algorithms

<https://doi.org/10.22214/ijraset.2022.39891>

1. Aldhyani, T.H.H.; Alzahrani,(2022) A. Framework for Predicting and Modeling Stock Market Prices Based on Deep Learning Algorithms. Electronics 2022, 11, 3149. <https://doi.org/10.3390/electronics11193149>
2. Aakash Agarwal (2021). Study of machine learning algorithms for potential stock trading strategy frameworks

<https://doi.org/10.35912/ijfam.v3i3.604>

1. Yogita Deshmukh1 , Deepmala Saratkar2 (2019). Stock Market Prediction Using Machine Learning

DOI 10.17148/IJARCCE.2019.8107

1. Rico Bayu Wiranata, Arif Djunaidy (2021). The Stock Exchange Prediction using Machine Learning Techniques: A Comprehensive and Systematic Literature Review
2. Memoona Shaheen, Mehreen Arshad (2020). Use of Machine Learning in Stock market Prediction
3. Grace Yoby Dopi, Rudy Hartanto (2021). Systematic Literature Review: Stock Price Prediction Using Machine Learning and Deep Learning.

<https://dx.doi.org/10.2991/aebmr.k.211117.008>