DAL 2023: Final Exam

Advait Amit Kisar
Dept. of Aerospace Engineering
Indian Institute of Technology Madras
Chennai, India
ae20b007@smail.iitm.ac.in

***Abstract*—This final exam deals with studying the share prices of different companies listed in the National Stock Exchange (NSE - India) and the NASDAQ (in the United States). It also deals with studying the USD-INR Exchange Rate. All this data is mostly between the years 2019 and 2021. It would help us to understand which shares perform in a bullish or bearish way and also would help to predict the future prices of these shares. For this analysis, I have used the Long Short-Term Memory (LSTM) on the training data to predict the present and future prices of the shares of these organizations**

***Index Terms*—Long Short Term Memory, Recurrent Neural Network, Moving Averages**

I. Introduction

This assignment studies the variation of share prices over time for different organizations and the USD-INR Exchange Rate. Analyzing this data will help us to understand the present patterns and predict the performance of these shares in the near future.

The Long Short-Term Memory (LSTM) is a special type of Recurrent Neural Network, which is specifically useful for sequential data. In conventional Neural Networks, a single data point is processed and the output is given by the model. However, in sequential data like Share Prices, Blood Pressure Variation with Time, and Natural Language Processing, the current data point's output not only depends on itself but also on the past data points. Hence, LSTM is a highly effective model that captures the past and present data to predict the output accurately in the case of sequential data, textual data, and time-series data.

Using the Long Short Term Memory, I have modeled the share prices as a function of the share prices in the past and the date. All the columns are numerical data, of which there are few missing values, which are filled and then LSTM is used to predict the share prices to generate insights about the performance of the share prices and this will help the investors to decide whether a particular stock is good enough for investing or not.

This paper deals with predicting the share prices of different organizations listed in NSE and NASDAQ, and also the USD-INR Exchange Rate. The model employed here is the Long Short Term Memory as we need to process time-series data which is sequential in nature. Thus, the Long Short-Term Memory is the best way to deal with Sequential Data from the class of Deep Learning Models like Neural Networks, GRU, etc. This analysis would help us in studying the performance of the share prices and help in predicting their future prices.

II. Long Short Term Memory

The Long Short Term Memory (LSTM) model is a Deep Learning Model used to train sequential data like Time Series Data, and Natural Language Processing. In sequential data, the target variable also depends on the present data point as well as the data points in the past. The LSTM model is an improved version of the Recurrent Neural Network. Recurrent Neural Networks (RNN) are a special class of Neural Networks that are used for training and modeling sequential data. In the case of Recurrent Neural Networks, they suffer from the problem of vanishing gradients. This problem has been solved using LSTM by incorporating three gates in a unit of LSTM.

A. Architecture of LSTM

The architecture of a Long Short Term Memory Unit is given below:-

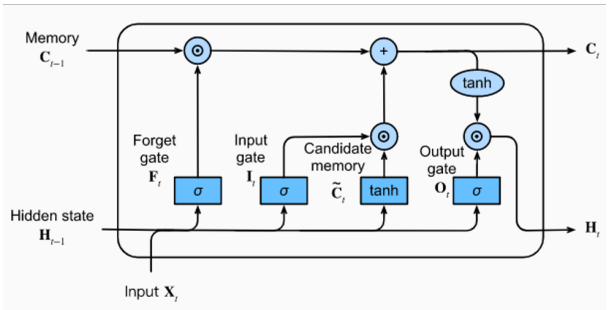

Fig. 1. Architecture of LSTM Unit

As shown above, an LSTM Unit has three gates, namely, Forget Gate, Input Gate, and Output Gate. These gates help to solve the problem of vanishing gradients because of long-term dependencies. The cell state or the Long Term Memory is denoted by C_t and the hidden state of the Short Term Memory is denoted by H_t for the t^{th} timestamp. As these two states help to flow the information in the model, this is called Long Short Term Memory. The functions of these gates are as follows:-

1) Forget Gate:-
   a) The job of Forget Gate is to decide whether to pass the information of previous steps or the past to the present or not.
   b) The Forget Gate's equation is given by:-

$$F_t = \sigma(W_f[H_{t-1}, X_t] + b_f) \tag{1}$$

where, W_f is a weight matrix, b_f is the bias term.

2) Input or Update Gate:-
   a) The job of Input Gate is to decide the importance of the information of the current timestamp.
   b) The Input Gate's equation is given by:-

$$I_t = \sigma(W_i[H_{t-1}, X_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(W_c[H_{t-1}, X_t] + b_c) \tag{3}$$

   where, W_i and W_c are weight matrices, and b_i and b_c are the bias terms.

3) Output Gate:-
   a) The job of the Output Gate is to generate the output for the LSTM Unit.
   b) The Output Gate's equation is given by:-

$$O_t = \sigma(W_o[H_{t-1}, X_t] + b_o) \tag{4}$$

   where, W_o is a weight matrix, b_o is the bias term.

Finally, the output is processed as follows:-

$$C_t = I_t\tilde{C}_t + F_tC_{t-1} \tag{5}$$

$$H_t = O_tC_t \tag{6}$$

In this way, LSTMs solve the problem of Long Term Dependencies and generate predictions on sequential data.

B. Working of LSTM

The working of the LSTM can be explained as follows:-

1) The data is split into training and test sets with an interval set for considering the previous timestamps.
2) Fit the LSTM model by suitably choosing an architecture and hyperparameters.
3) After the training phase, use the test set on this SVM to generate predictions.

C. Advantages and Disadvantages of LSTM

Long Short-Term Memory is a highly effective RNN model for solving problems of vanishing or exploding gradients because of long-term dependencies. It is best suited for long sequences of data and handles the gradients well. The only disadvantage of this algorithm is that it requires more time during the training phase as it is fairly complex compared to basic RNN models.

III. Analysis and Observations

The problem was to understand the variation of share prices of different organizations and generate predictions on given data. Using this, we can understand which shares are profitable and this will help the investors and traders to choose the right shares they would like to invest to maximize their profits. The dataset was cleaned and modified to make it consistent enough to train a model. The share prices were visualized to understand their variation in a preliminary stage and Moving Averages were plotted to understand their performance on longer durations. The data was normalized based on its distribution to fit the model to enhance its prediction capability. The model being used is the Long Short-Term Memory. All these points have been discussed in detail in the following subsections.

A. Problem and Data Visualization

The problem is to understand and model the share prices' variation for different organizations listed in NSE and NASDAQ, and also the USD-INR Exchange Rate.

There are 7 datasets namely, Cognizant, HCL Technologies, HDFC Bank, ICICI Bank, Infosys, SBI Bank, and USD-INR Exchange Rate. These datasets differ in number of samples and their dates. However, most of these datasets have the share prices from 2019 to 2021. Also, the 6 organizations' data has the same 6 columns, namely, Date, Open, High, Low, Close, and Volume of Share Prices. In the case of the USD-INR Exchange Rate, the column Adjusted Close is in place of Volume. The data in the Cognizant dataset is in USD, and the rest of the datasets have their data in INR.

When the datasets were analyzed, there was a significant amount of missing values in many columns in most of the datasets. The most common way of dealing with missing values is mean imputation, but this was not useful in this particular case, as the missing values were significantly higher. If these had been imputed by the mean value, the variance of the normal distribution would drop and the exponential distribution would not remain purely exponential. However, these methods do not work for sequential data like share prices. Hence, I have used linear interpolation to fill the missing values so that the nature of the variation of share prices does not change much.

All the indicators have more or less similar variations for a given company. Hence, only the Opening Prices of shares are plotted below to get an overview of the variation of these indicators for the different organizations mentioned above. In addition, the plot of HDFC Bank's columns are plotted below to visualize all columns of the data.

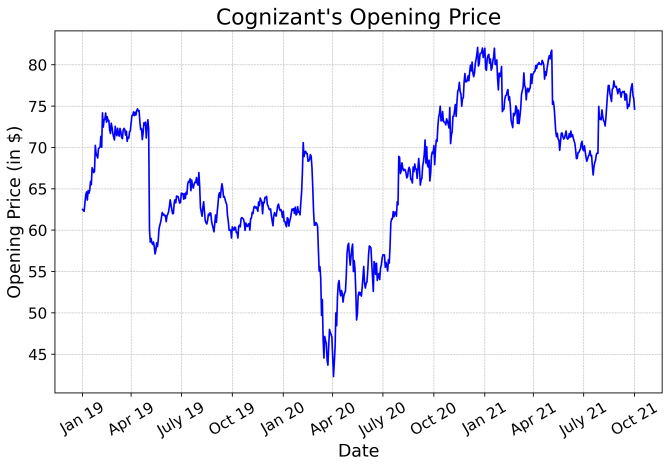

Fig. 2. Cognizant's Opening Price from 2019 to 2021

Fig. 3. HCL Technologies' Opening Price from 2019 to 2021

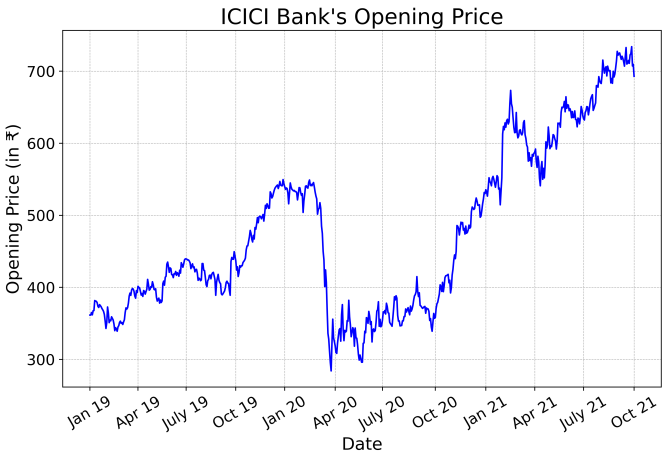

Fig. 4. ICICI Bank's Opening Price from 2019 to 2021

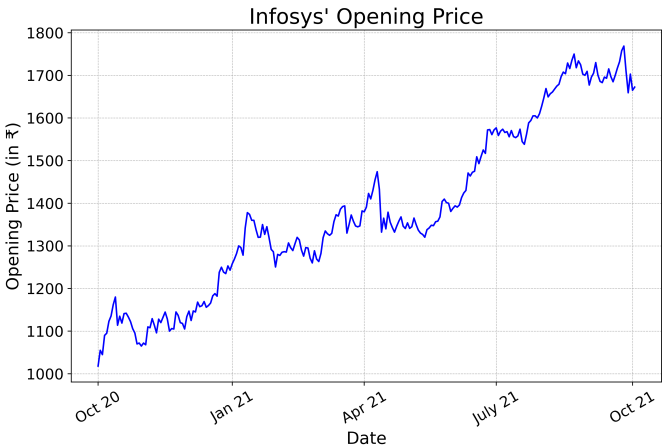

Fig. 5. Infosys' Opening Price from 2019 to 2021

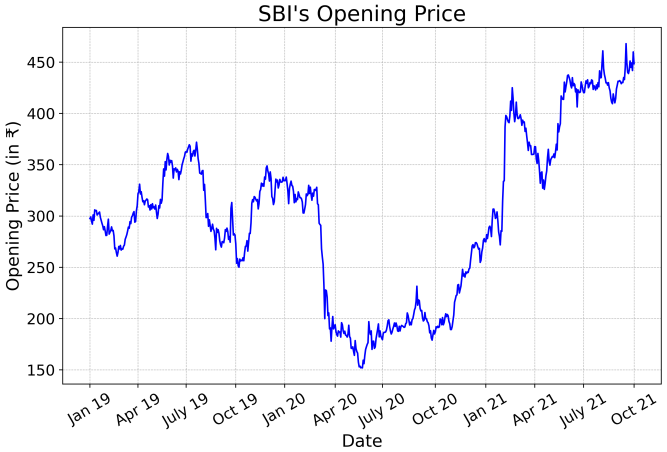

Fig. 6. SBI Bank's Opening Price from 2019 to 2021

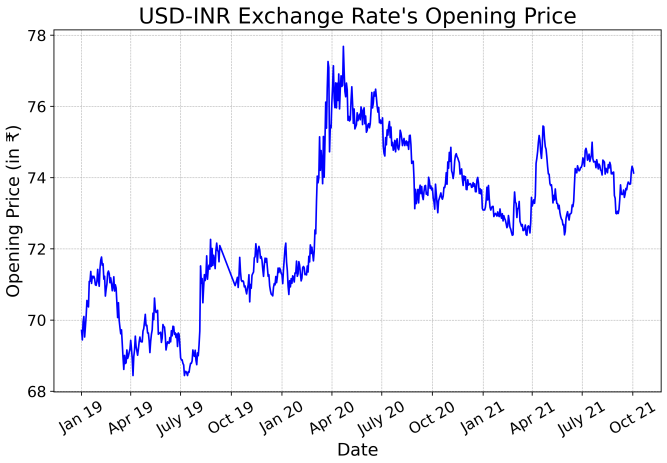

Fig. 7. USD-INR Exchange Rate's Opening Price from 2019 to 2021

The plots for HDFC Bank's columns, namely, Opening, High, Low, Closing Prices and Volume are plotted below:-

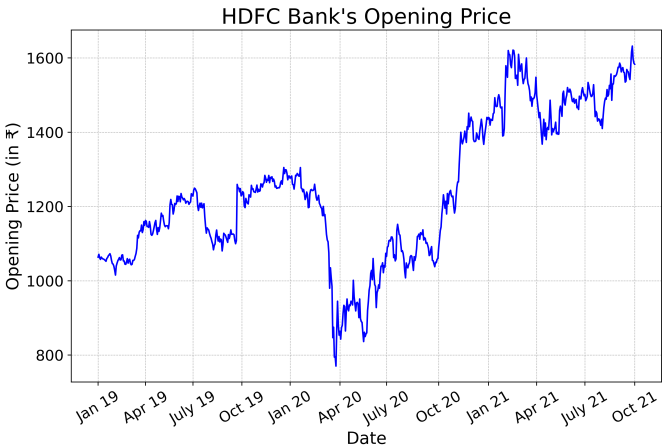

Fig. 8. HDFC Bank's Opening Price from 2019 to 2021

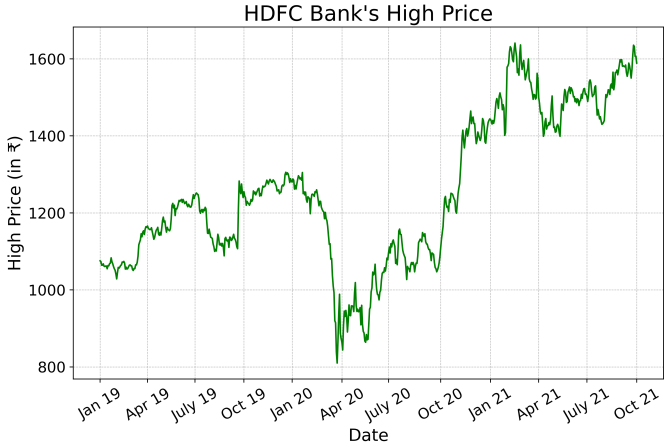

Fig. 9. HDFC Bank's High Price from 2019 to 2021

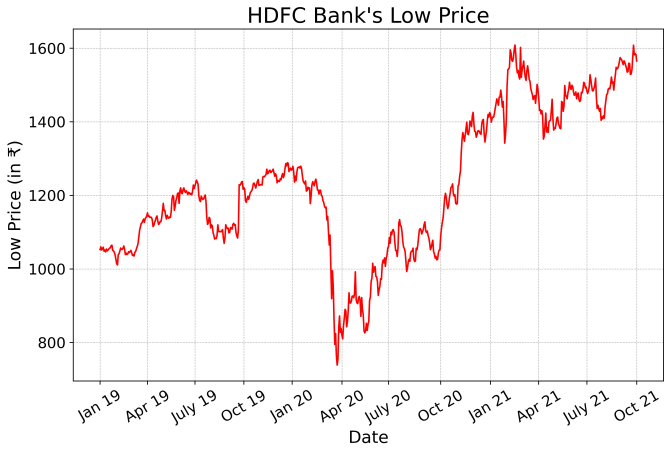

Fig. 10. HDFC Bank's Low Price from 2019 to 2021

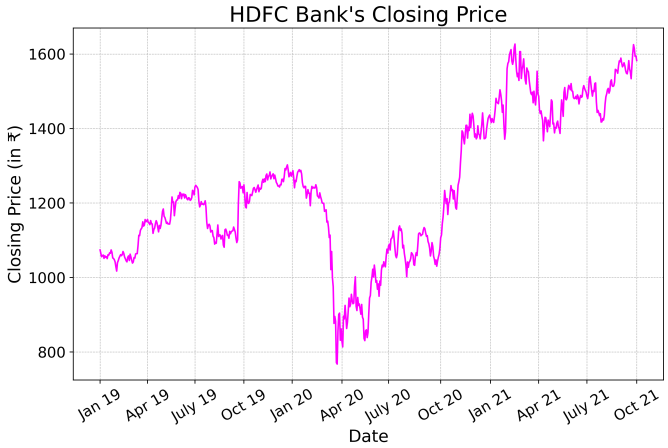

Fig. 11. HDFC Bank's Closing Price from 2019 to 2021

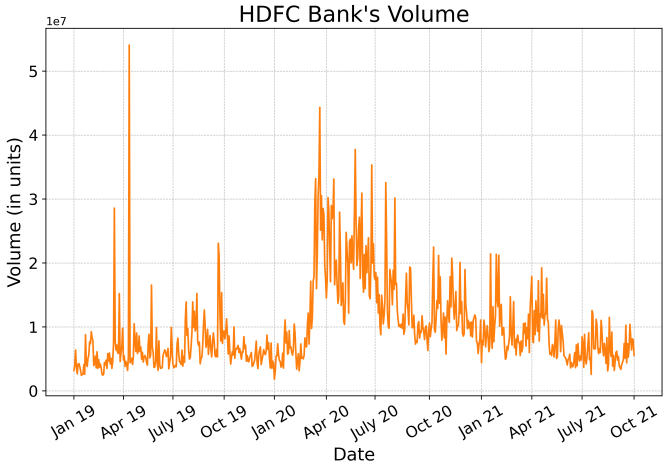

Fig. 12. HDFC Bank's Volume from 2019 to 2021

B. *Moving Averages*

Moving Averages are a type of technique used to visualize the performance of a stock over time using the average over a particular duration. The Moving Averages of HDFC Bank and USD-INR Exchange Rates are plotted below:-

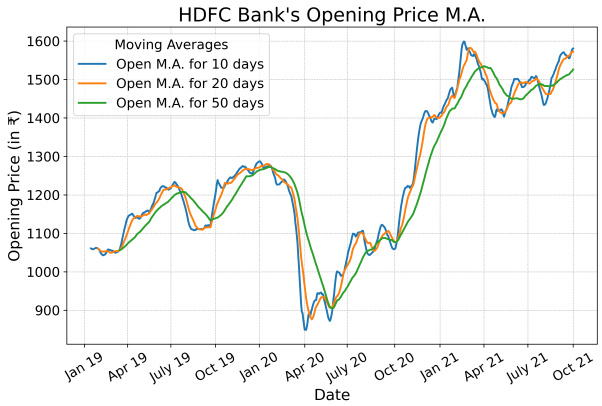

Fig. 13. HDFC Bank's Opening Price Moving Average from 2019 to 2021

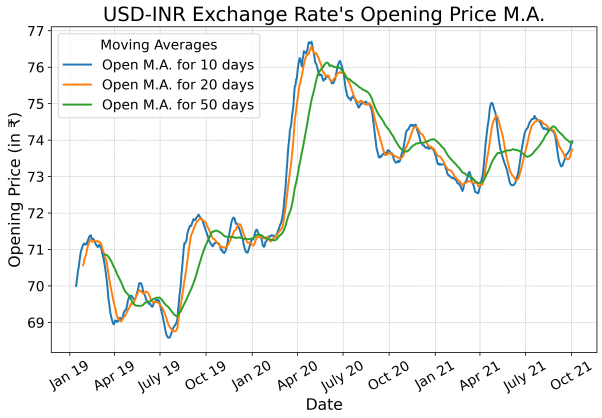

Fig. 14. USD-INR Exchange Rate's Opening Price Moving Average from 2019 to 2021

C. Working of the Model

The steps involved in making the model are discussed in this section. The model being used is the Long Short-Term Memory.

The next step is to normalize the data in the input features. This ensures that the data lies between 0 and 1. This is performed by using 'MinMaxScaler()' from Scikit-Learn.

Each of the datasets is split into training and test sets in an 80:20 ratio, with an interval of 60 days for taking the inputs from past data. This data is rearranged in an array in which each element has a length of 60 days.

This data is modeled using LSTM from Keras, and the number of epochs for most of the columns and datasets is 50. For a few columns, the number of epochs was increased to 100. The architecture used for the LSTM Model consists of two layers of LSTM units, which have 128 and 64 units respectively. For columns other than Volume, additional two dropout layers are added with a dropout value of 0.2. Two fully connected layers of 25 neurons and 1 neuron are followed by these LSTM layers.

For Infosys and HCL Technologies, the above model was not performing well, hence another architecture was used. In this model, there was a single LSTM layer of 100 units with a dropout layer with a dropout value of 0.2 if the column is not of Volume. This is followed by a fully connected layer of 1 neuron which returns the output.

The MAE for the test set of all the columns and datasets are tabulated below. The units for the share price of Cognizant is USD and the rest is INR. The unit for Volume is units. The USD-INR Exchange Rate has an Adjusted Close Column while the rest of the datasets have Volume.

Dataset	Open	High	Low	Close	Vol./Adj. Close
Cognizant	0.83	0.86	0.83	0.89	9.55×10^5
HCL	52.89	37.94	43.55	40.88	1.81×10^6
HDFC	19.94	18.53	18.38	19.71	2.25×10^6
ICICI	13.36	9.37	9.95	9.60	7.20×10^6
Infosys	25.18	26.05	17.91	19.39	1.27×10^6
SBI	6.51	6.06	7.12	6.91	1.23×10^6
USD-INR	0.24	0.24	0.20	0.24	0.24

TABLE I
MEAN ABSOLUTE ERROR OF ALL COLUMNS AND DATASETS

The reason for the spike in error for the Volume of the datasets is that there are numerous spikes in between which are difficult to capture using LSTM.

D. Insights and Observations

The predictions made using the LSTM Model are plotted below for Opening Prices of different datasets as follows:-

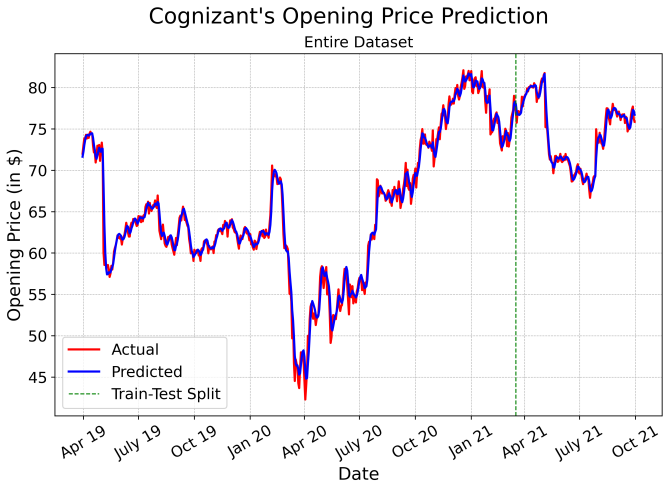

Fig. 15. Prediction for Cognizant's Opening Price

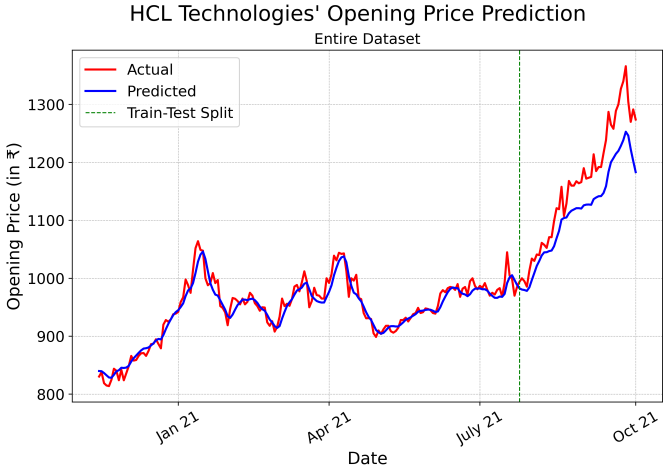

Fig. 16. Prediction for HCL Technologies' Opening Price

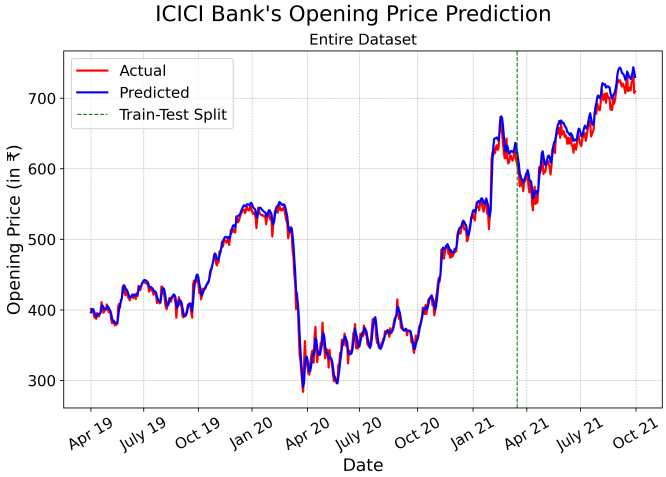

Fig. 17. Prediction for ICICI Bank's Opening Price

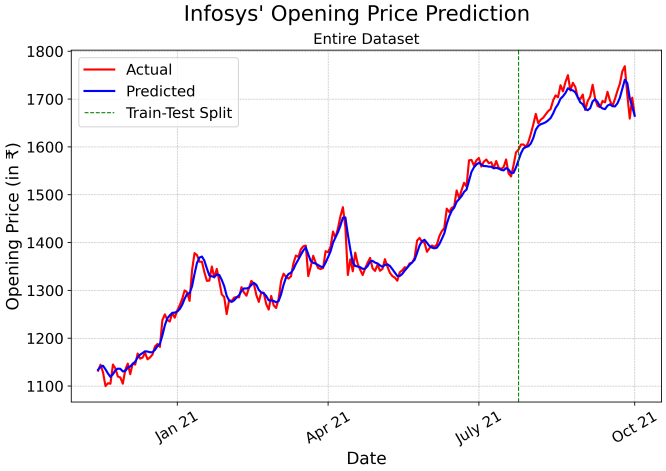

Fig. 18. Prediction for Infosys' Opening Price

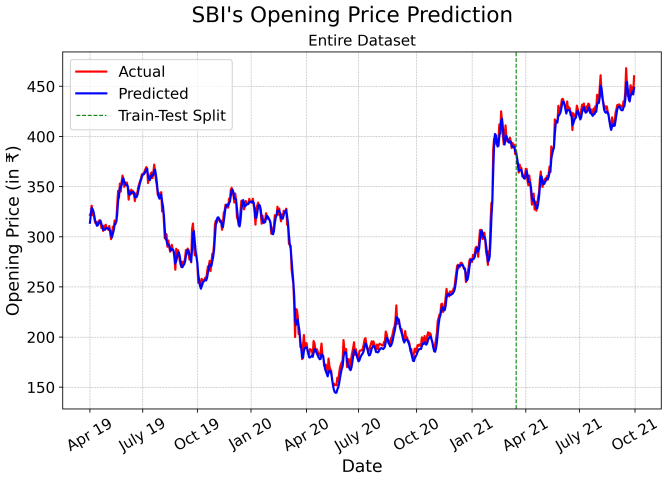

Fig. 19. Prediction for SBI Bank's Opening Price

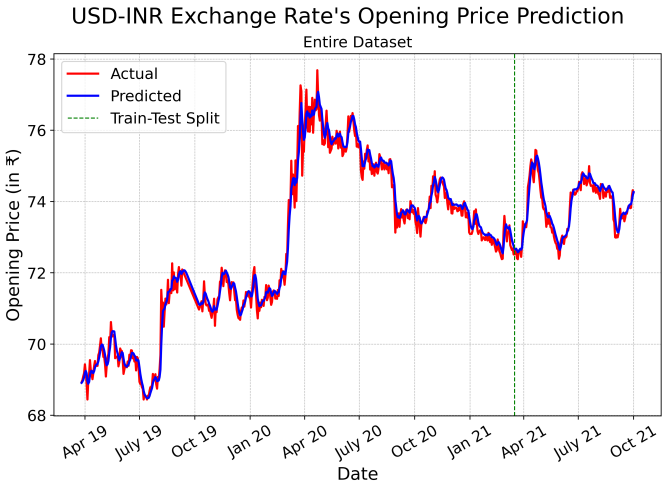

Fig. 20. Prediction for Exchange Rate's Opening Price

It can be observed that the LSTM models are highly accurate in predicting the Opening Price in most cases. The accuracy was best when the train-test ratio was set to 80:20. For ratios higher than this, it does not work well, as the data available for training is different than the test data. Also, dropout regularization was used so that the model does not overfit on the past data.

The variation in price indicators are similar, hence, only the Opening Price has been plotted. The best fit for price indicators is obtained for the USD-INR Exchange Rate and the worst for HCL Technologies, depending on the Mean Absolute Error. The predictions for HDFC Bank's Opening Price on the entire dataset and test dataset are plotted below:-

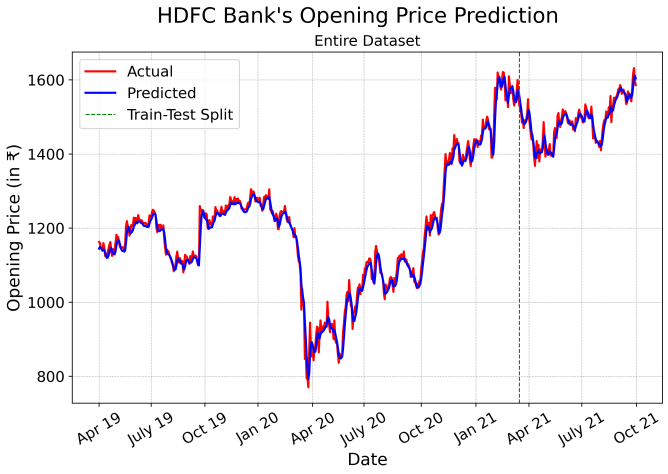

Fig. 21. Prediction for HDFC Bank's Opening Price (Entire Dataset)

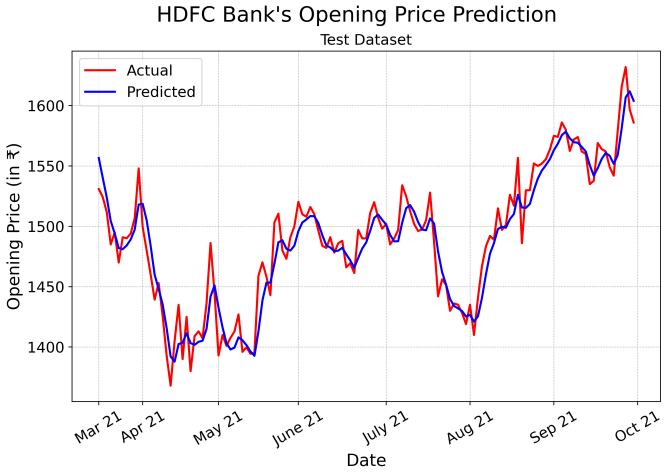

Fig. 22. Prediction for HDFC Bank's Opening Price (Test Dataset)

The predictions for HDFC Bank's Volume on the entire dataset and test dataset are plotted below:-

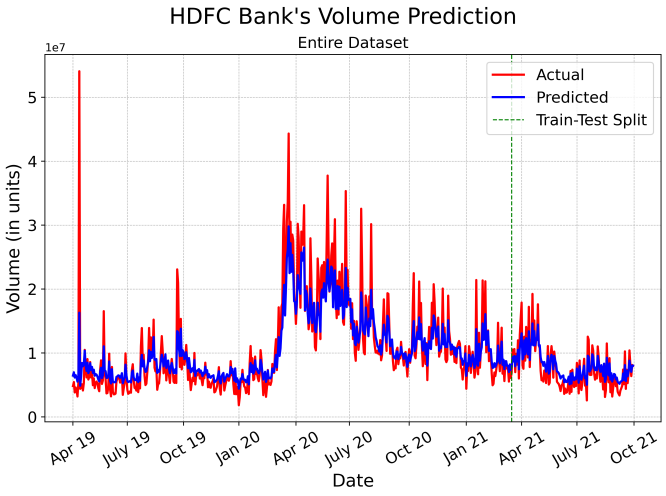

Fig. 23. Prediction for HDFC Bank's Volume (Entire Dataset)

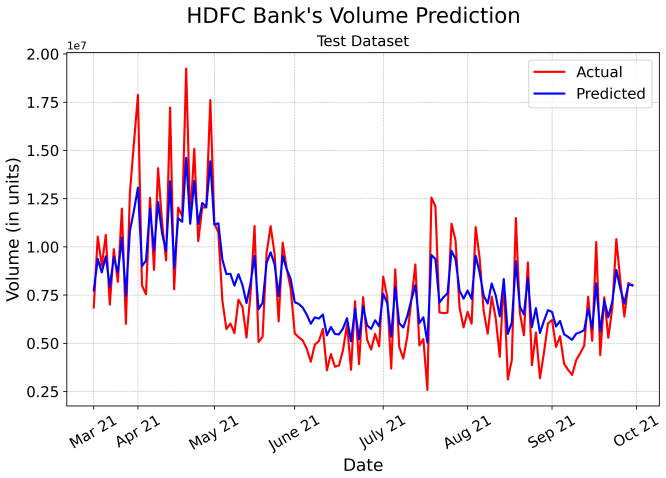

Fig. 24. Prediction for HDFC Bank's Volume (Test Dataset)

IV. Conclusion

To summarize all the observations made so far using the predictions we get from the datasets, the Volume column has a lot more sudden spikes as compared to share prices. Hence, it is difficult to model the Volume column as compared to price indicators. The best fit for price indicators is obtained for the USD-INR Exchange Rate and the worst for HCL Technologies, depending on the Mean Absolute Error.

Further analysis could include using better algorithms Echo State Networks (ESN), Ensemble Methods, Attention Mechanisms, and Transformer Networks. Hybrid Approaches can also be used along with LSTMs to improve prediction capability. Deep Reinforcement Learning can also be used to suggest to the traders the right action to execute so that they can maximize their profit. Graph Neural Networks (GNNs) can also be used to capture interactions between different stocks.

References

[1] Implementation of LSTM in Keras, [documentation]
[2] What is LSTM?, [Analytics Vidhya]