Stock Price Prediction using LSTM Model

Abstract:

It has never been easy to invest in a set of assets, the ups and downs of financial market does not allow simple models to predict future asset values with higher accuracy. Machine learning, which consists of making computers perform tasks that normally requiring human intelligence is currently the dominant trend in scientific research. This article aims to build a model using Recurrent Neural Networks (RNN) and especially Long-Short Term Memory model (LSTM) to predict future stock market values.

Problem Statement:

Stock market is a promising financial investment that can generate great wealth. However, the volatile nature of the stock market makes it a very high risk investment. Stock market prediction has always been an important issue in the field of technology, financial and mathematics because of its very likely financial gain. So to reduce the burden of analyzing all the previous year’s stocks manually we came up with a solution which could predict the stock prices based on the previous trends.

Methodology:

Initially we are retrieving stock data from yfinance module. The data is scaled using the MinMaxScaler from sklearn module. Now 80% of the data is taken for training the LSTM model. LSTM (Long short-term memory) is an artificial recurrent neural network architecture used in the field of deep learning. LSTMs are very powerful in sequence prediction problems because they’re able to store past information. The LSTM model can be tuned for various parameters such as changing the number of LSTM layers, adding dropout value or increasing the number of epochs. After the training the predictions made by the model are plotted on a graph using matplotlib. In which we can identify the difference between the actual closing price and the predictions.

Software and Hardware Used:

Software:

Programming language: Python

Machine Learning Packages:

1. TensorFlow

2. Keras

3. pandas

4. Numpy

5. sklearn

6. yfinance

7. matplotlib

8. BeautifulSoup

Hardware:

Executed on Laptop (i5 8 th generation, 8GB RAM)

Technical Description:

1) Using the yfinance module we will retrieve the yfinance codes corresponding to the companies from the website <https://stockanalysis.com/stocks/> and searches for the given parameter (i.e. company).

2) In the dataframe function, the DataFrame is created using the Ticker function from yfinanace module. The data is cleaned in accordance with the Requirements.

3) Now cleaned Data is scaled and each feature is translated individually such that it is in the given range on the training set between zero and one. Later, we create the x\_train\_data and y\_train\_data as numpy arrays which will be used for training the model.

4) We are training the lstm model with the training data.

5) Now the model predicts the values of the testing data and those are plotted in a graph using matplotlib.

Model used:

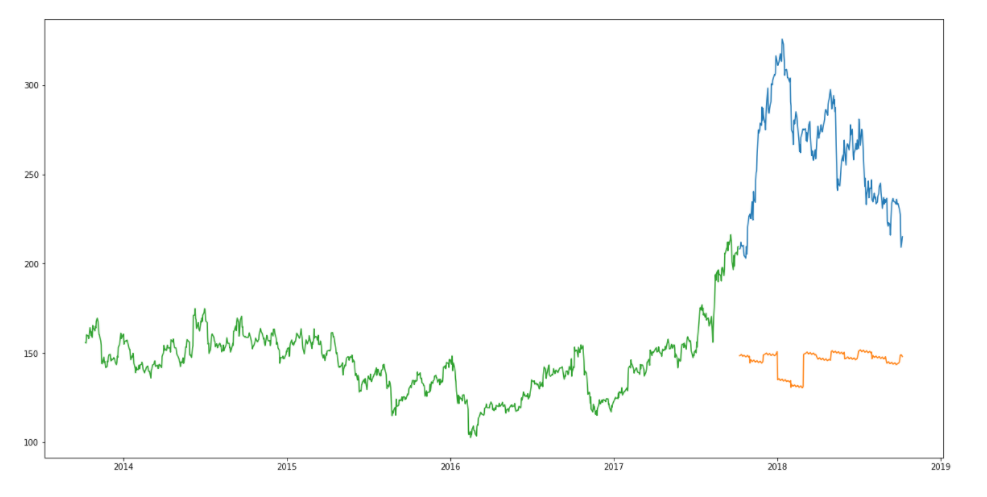
LSTM (Long short-term memory) is an artificial recurrent neural network architecture used in the field of deep learning. LSTMs are very powerful in sequence prediction problems because they’re able to store past information. This is important in our case because the previous price of a stock is crucial in predicting their future price. We have some other methods for predicting the prices like

1) Linear Regression

2) K Nearest Neighbors

Linear Regression:

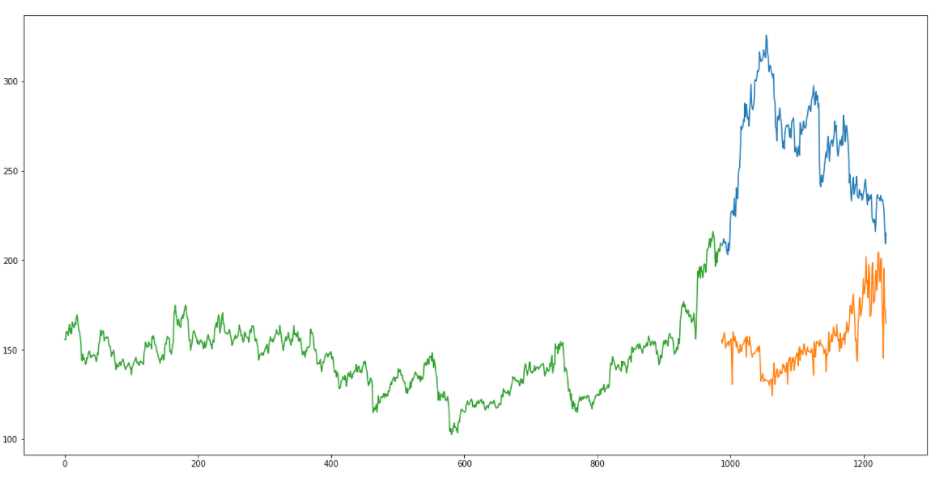
Linear regression is a simple technique and quite easy to interpret, but there are a few obvious disadvantages. One problem in using regression algorithms is that the model overfits to the date and month column. Instead of taking into account the previous values from the point of prediction, the model will consider the value from the same date a month ago, or the same date/month a year ago.



Predictions on Tata Global stocks using Linear Regression Model

K Nearest Neighbours:

The RMSE value is almost similar to the linear regression model and the plot shows the same pattern. Like linear regression, kNN also identified a drop in January 2018 since that has been the pattern for the past years. We can safely say that regression algorithms have not performed well on this dataset.

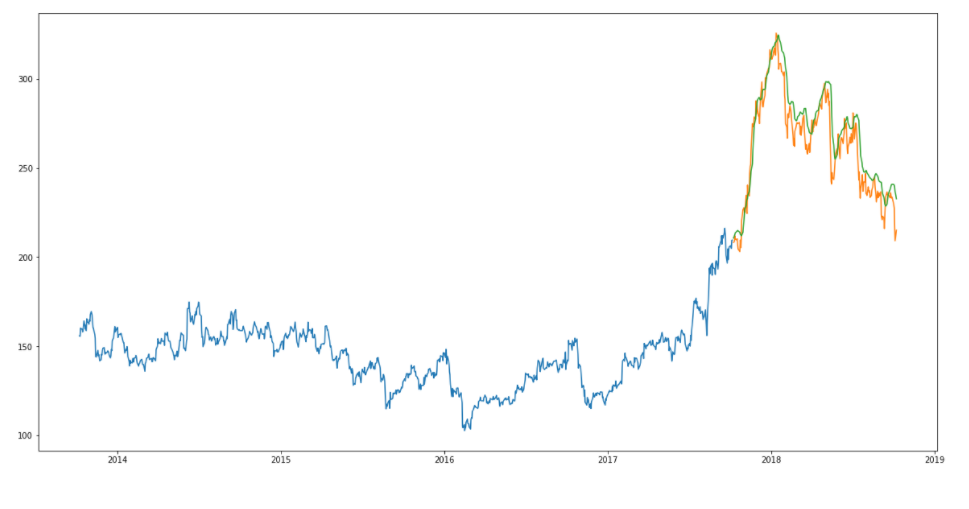


Predictions on Tata Global stocks using kNN Model

LSTM Model:

LSTMs are widely used for sequence prediction problems and have proven to be extremely effective. The reason they work so well is because LSTM is able to store past information that is important, and forget the information that is not. LSTM has three gates:

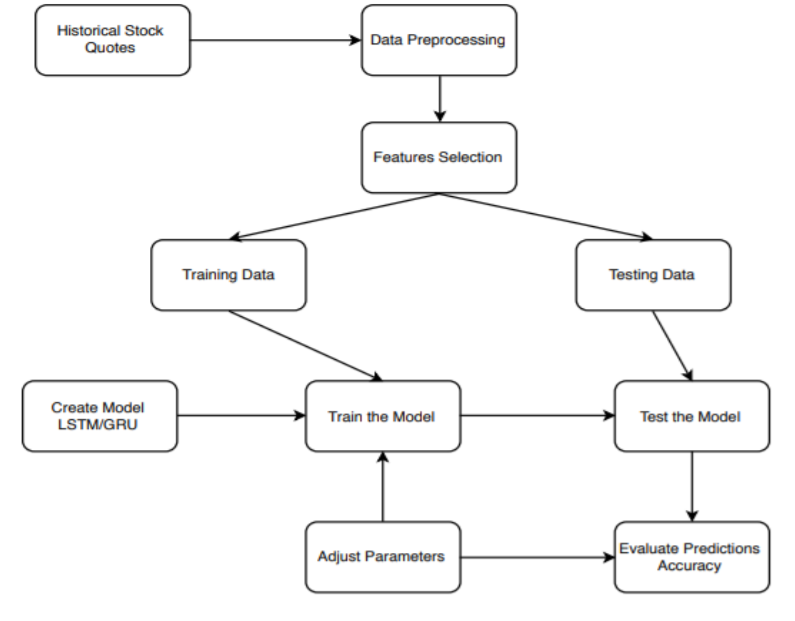
* **The input gate:** The input gate adds information to the cell state
* **The forget gate:** It removes the information that is no longer required by the model
* **The output gate:**Output Gate at LSTM selects the information to be shown as output



Predictions on Tata Global stocks using LSTM Model

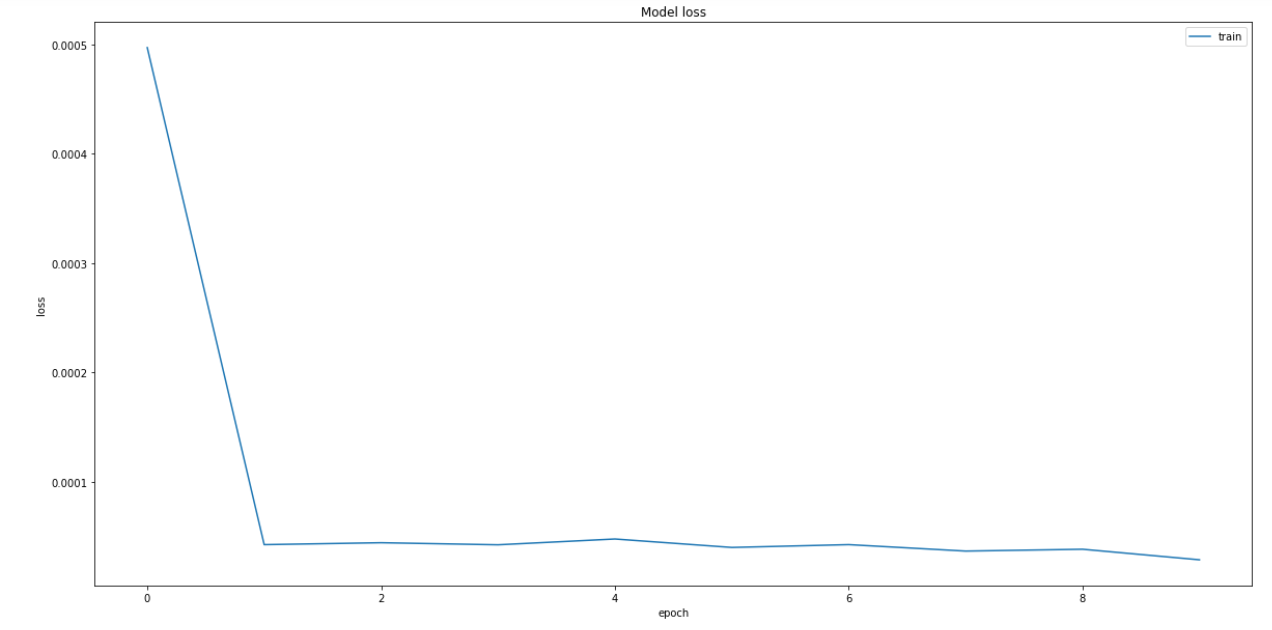
Source Code:

Process Diagram:



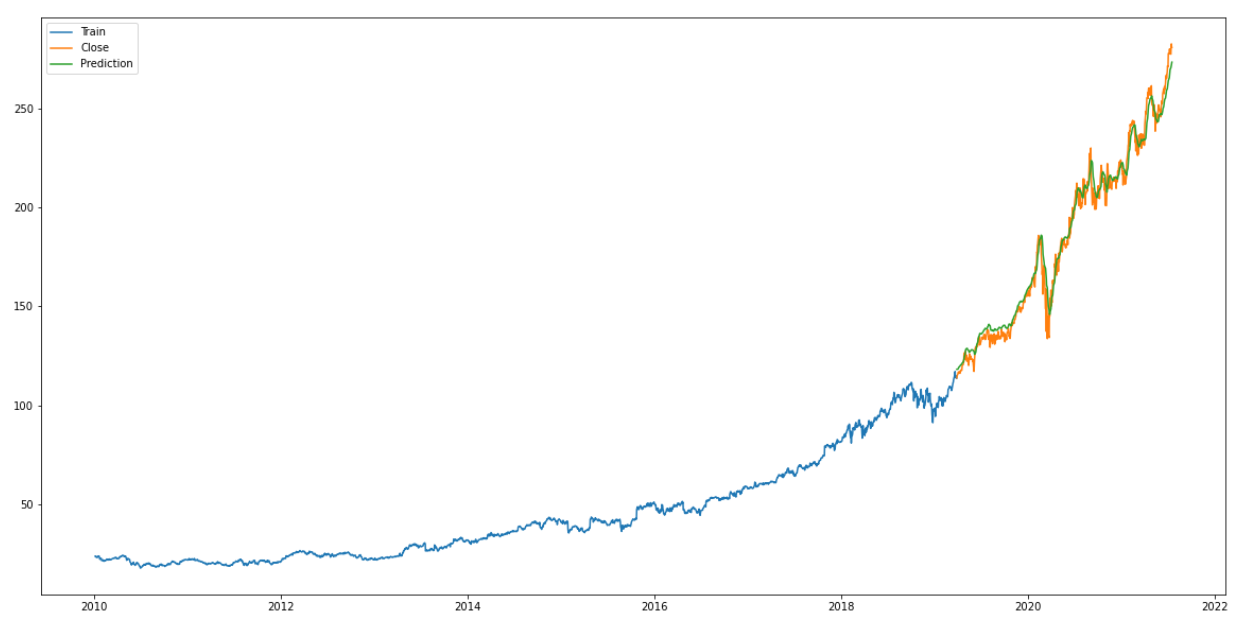
Performance:

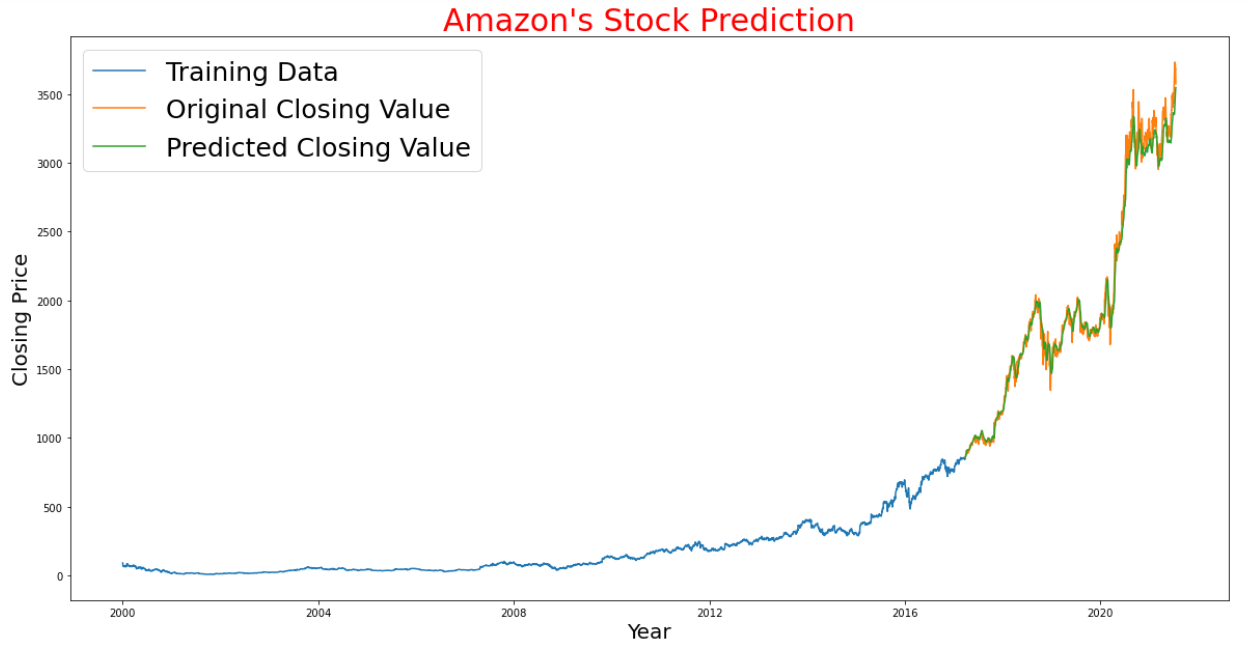
Training loss on Amazon Stock Data



Result:

Microsoft Stock Predictions





By:

Team number: 23

A.Yamini -198W1A0569

G.S.S.V Badrish -198W1A0580

K.G.N. Prabhanjali -198W1A0585

T.Avinash -198W1A05C0