

Stock Market Price Trend Prediction

Mentor : Prof. M. V. Joshi

Group Members

201801131: Yash Shah

201801157: Keval Thakarar

201801168: Param Modi

201801454: Jaymin Chauhan

Table of Content :

- 1) [Introduction](#)
- 2) [Problem Statement](#)
- 3) [Datasets](#)
- 4) [Methodology](#)
 - a) [Raw Data](#)
 - b) [Feature Expansion](#)
 - c) [News Sentiments & Google Trend analysis](#)
 - d) [LSTM](#)
- 5) [References](#)
- 6) [Results](#)

Introduction

The main aim of our project is short term price / trend prediction of stock. We have implemented our algorithm based on the algorithm given in the Research Paper “Short term stock market price trend prediction using comprehensive deep learning” by J.Shen and M.Omar Shafiq with some changes and tried to predict the Close Price of the stock for a particular day.

Motivation

Stock market is the place where people can make a fortune if they are successfully able to predict the price trend of the stock market. It's kind of impossible to predict the price trend of the stock price accurately because it depends on various parameters like economic condition, operator's mood, company's performance and also on the recent news. There are financial institutions and brokers who estimate the price trend of the stock market and give the targets, so why can't a machine predict the price trend?, How accurate are they in predicting the price trend?, Can they predict the price accurately?. Hence this topic is a hot favourite among the financial and technical researchers. So, we decided to choose this as our project topic.

Problem Statement

Implement an algorithm to predict the “Close Price” of the stock for a particular day based on the Past Performance, Market Indicators, Technical Indicators, Trend, News Sentiments and Performance of other Companies.

Datasets

S&P 500 Companies

To get the list of Companies included in the S&P 500 List, We have collected this data from the Wikipedia Page using Web-Scraping.

```
import bs4 as bs
import pickle
import requests

def save_tickers():
    resp=requests.get('https://en.wikipedia.org/wiki/List_of_S%26P_500_companies')
    soup=bs.BeautifulSoup(resp.text)
    table=soup.find('table',{'class':'wikitable sortable'})
    tickers=[]
    for row in table.findAll('tr')[1:]:
        ticker=row.findAll('td')[0].text[:-1]
        tickers.append(ticker)

    with open("tickers.pickle",'wb') as f:
        pickle.dump(tickers, f)

    return tickers

save_tickers()
```

```
['MMM',
 'ABT',
 'ABBV',
 'ABMD',
 'ACN',
 'ATVI',
 'ADBE',
 'AMD',
 'AAP',
 'AES',
 'AFL',
 'A',
 'APD',
```

YFinance:

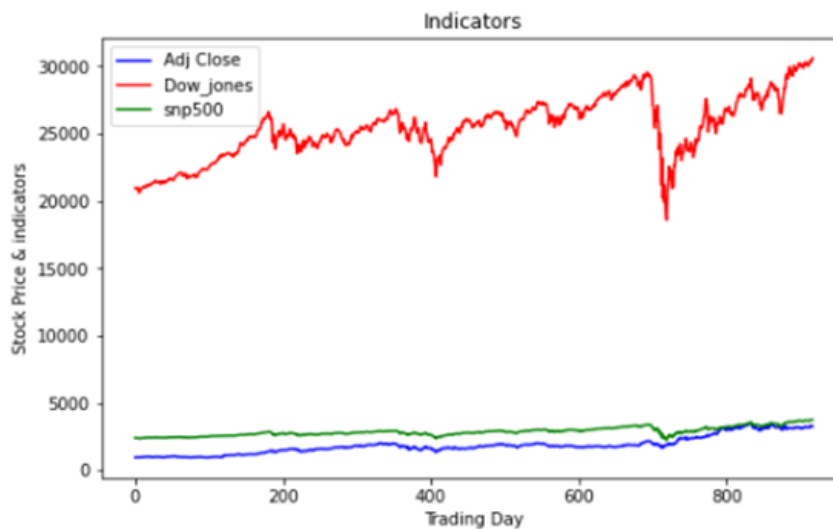
We have used the yahoo finance library to get the data for the S&P 500 Company List obtained from WikiPedia. Here, We have used the download function with 3 parameters : Name of the Company, Start Date and End Date. We obtained the data for companies from 10-05-2017 to 31-12-2020. Which gives us information like High, Low, Open, Close, Adj. Close and Volume for that company on a particular day.

```
#Reading CSV file generated by our code
df_target=pd.read_csv('/content/drive/MyDrive/ML Lab Data/target checkpoint/dataset_target.csv')
df_target.head()
```

	Date	High	Low	Open	Close	Volume	Adj Close
0	2017-05-10	953.750000	945.000000	953.500000	948.950012	2096500	948.950012
1	2017-05-11	950.289978	940.780029	945.109985	947.619995	2194100	947.619995
2	2017-05-12	962.789978	951.530029	954.500000	961.349976	3625900	961.349976
3	2017-05-15	963.150024	956.059998	958.729980	957.969971	4270600	957.969971
4	2017-05-16	970.059998	960.909973	961.000000	966.070007	3126100	966.070007

Index Data (Dow Jones, S&P 500)

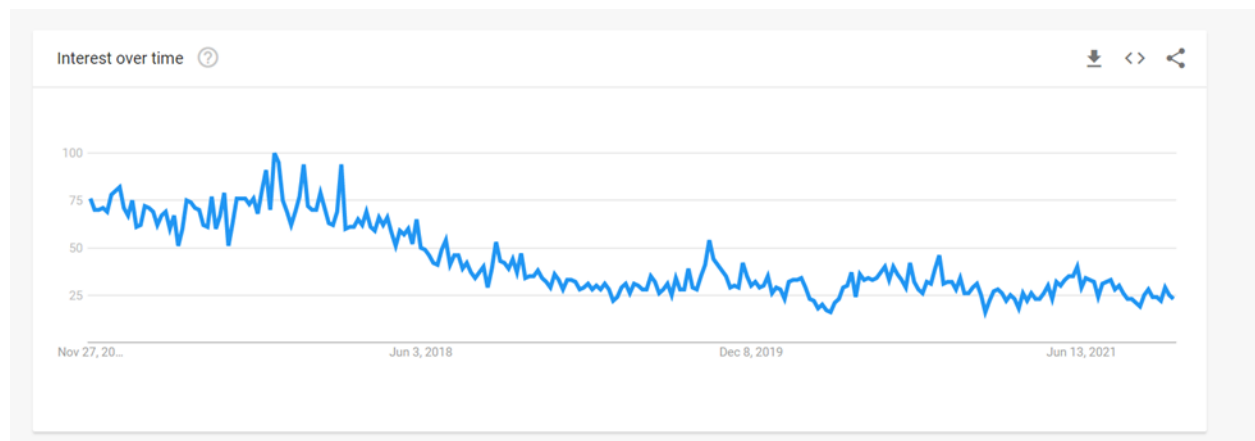
Indexes like Dow Jones and S&P 500 provide the sentiment of the whole Stock Market. Stock price also depends on the change in the Index. So, to make more accurate predictions and to consider Stock Market Sentiment, We have downloaded this data from [Macrotrends](#).



Google Trends

Stock Price is also dependent on the public sentiments. Google Trends is a website that evaluates the popularity of top Google Search queries in different

countries and languages. We have downloaded this data from Google Trends, Which gives Trends per week.



News Sentiments

News plays an important role in the Stock Price Trends. Because it creates sentiment about that particular Company like If a Company has made a Deal or Launched a new product or Posted good Results, Then this may shoot up the price of that company's stock. We have downloaded financial news from this [Site](#).

	News	Date
0	Amazon Signals Shift In 2021 Focus Following P...	2020-12-29
1	Amazon: How Amazon Wins	2020-12-27
2	Amazon - How I Handle This Company As A DGR In...	2020-12-23
3	Amazon Fresh Stores Are Going After Aldi, Not ...	2020-12-16
4	Why A \$4,000 Price Tag For Amazon Is Realistic	2020-12-14

Sources of Dataset :

- 1) [S&P 500 Companies](#)
- 2) [Yfinance](#)
- 3) [S&P 500](#), [Dow Jones](#)
- 4) [Google Trends](#)
- 5) [News Sentiments](#)

Methodology

Block Diagram



Fig 1 : Block Diagram

As shown in the above diagram, Our implementation can be divided into 5 major blocks.

1) Raw Data :

As explained above, We downloaded the Data Sets required for our implementation from above mentioned sites.

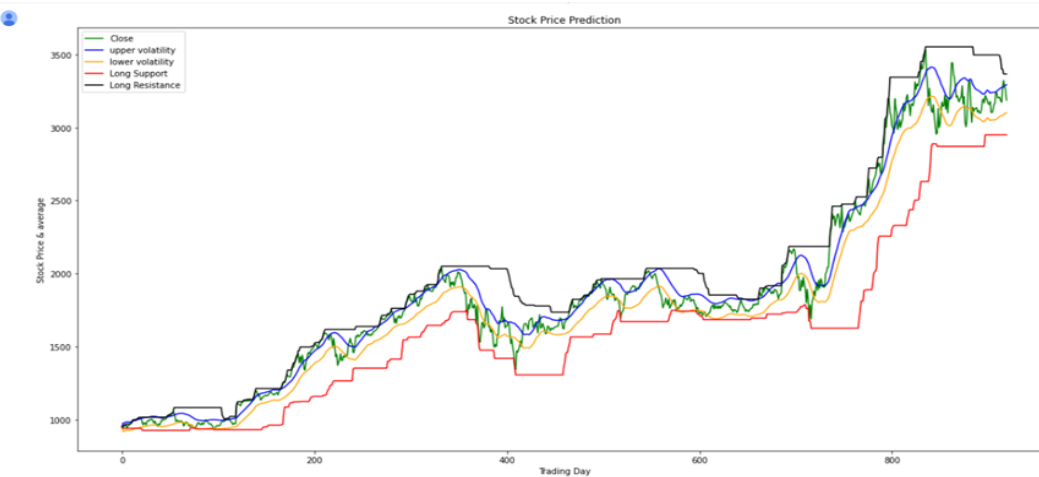
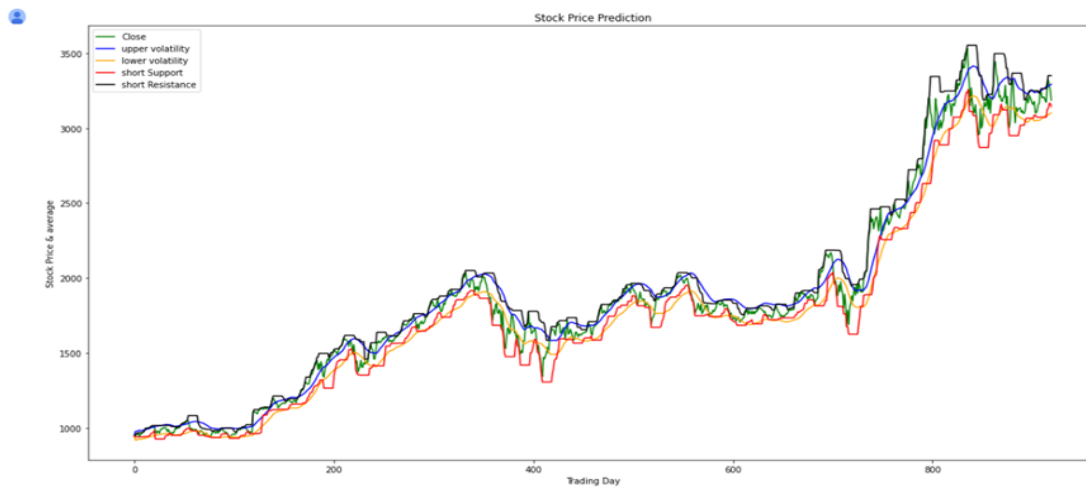
2) Feature Expansion :

In this stage, we have added some of the popular Technical Indicators for stock price prediction in our Dataset like Moving Average, Volatility Indexes, Support and Resistance.

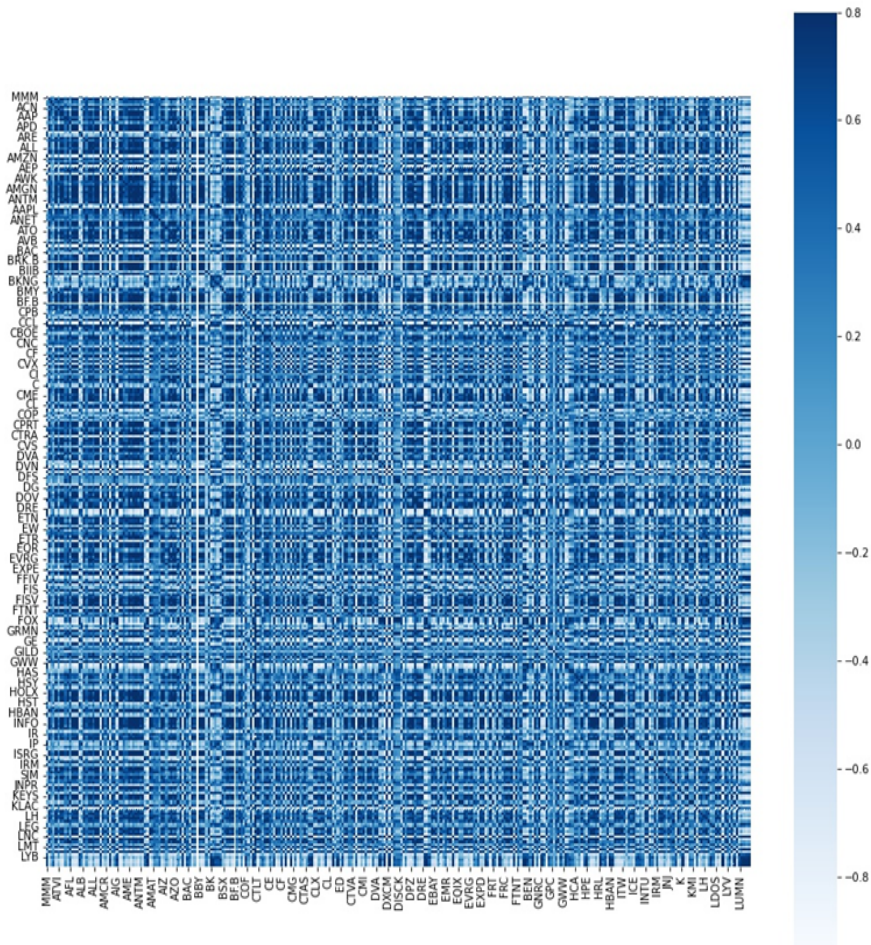
- a) Moving Averages :** Moving Average is a calculation used to analyze data points by creating a series of averages of different subsets of the full data set. We will calculate the 20 day Moving Average. This is popular among a lot of Traders for Stock Predictions.
- b) Support :** Support is a price level that functions as a floor, keeping the price from falling any further. Basically, there are buyers at that level who buy and push the markets back.
- c) Resistance :** Resistance, like support, is a price level that functions as a ceiling, preventing the price from rising any further. At that point, the selling basically chips in and pulls the stock price down.

There are two types of support and resistance based on the time limit. Long and Short.

d) Volatility Indexes : Expectations are reflected in volatility indexes. Based on the present time frame behaviour, two volatility indexes define a range or a band between which the predicted stock value for the following time frame may fall. The volatility indexes are normally established in relation to the moving average. The upper index is 2 percent of moving average + current moving average, and the lower index is current moving average – 2 percent of moving average, according to conventional procedure. As a result, a band is defined by this. When this band is broken, it indicates that huge volumes of transactions are taking place.



After adding some of the popular Technical Indicators, We checked How S&P 500 company's Stock price affects our Target Company's Stock price. For that, We checked Correlation between them. And took the top 10 Companies that have the highest correlation with our Target Company (Amazon).



3) News Sentiments & Google Trend Analysis :

For Google Trends data, We got data from Google Trends on a weekly basis but as our requirement is for daily basis, We converted the data to daily basis and added it as a feature to our main Dataset.

As stated in the Dataset Part, We downloaded the News for our target company Amazon from the Financial News website. To add it as a feature we need to know the Sentiment, Basically the Polarity (Positive or Negative) and

Strength (How much Positive?). For this we have used `SentimentIntensityAnalyzer()` of VADER Model.

The sentimental analysis of VADER is based on a lexicon that maps lexical elements to emotion intensities, which are referred to as sentiment scores. A text's sentiment score can be calculated by adding the intensity of each word in the text.

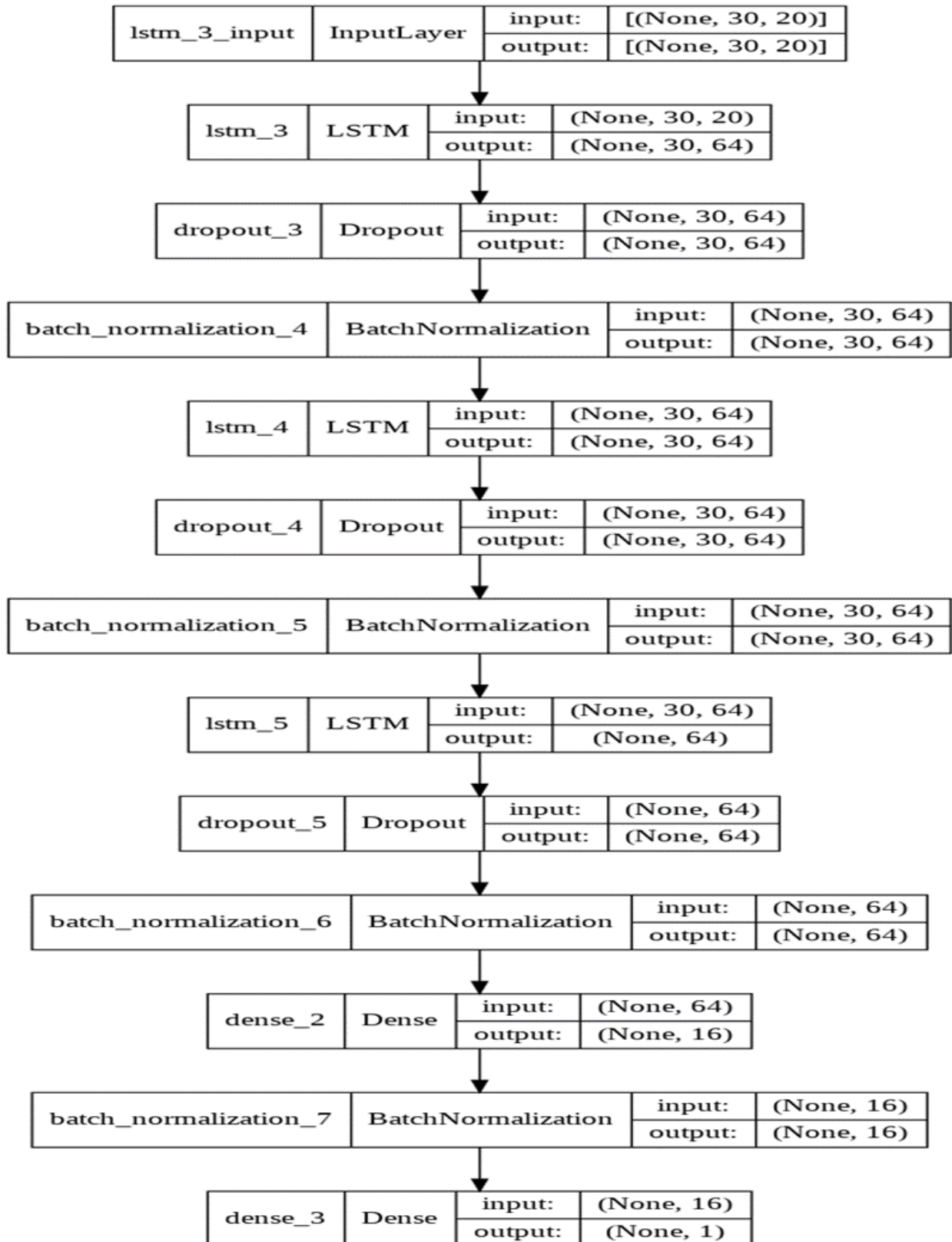
`SentimentIntensityAnalyzer()` is a VADER function that takes a string and produces a dictionary of scores in four categories:

- 1) Negative
- 2) Neutral
- 3) Positive
- 4) Compound

	News	Date	neg	neu	pos	compound
0	Amazon Signals Shift In 2021 Focus Following P...	2020-12-29	0.000	1.000	0.000	0.0000
1	Amazon: How Amazon Wins	2020-12-27	0.307	0.429	0.264	-0.1779
2	Amazon - How I Handle This Company As A DGR In...	2020-12-23	0.000	0.746	0.254	0.1779
3	Amazon Fresh Stores Are Going After Aldi, Not ...	2020-12-16	0.277	0.319	0.404	0.2960
4	Why A \$4,000 Price Tag For Amazon Is Realistic	2020-12-14	0.000	1.000	0.000	0.0000

Compound Score can be used to label the news as Positive (Score ≥ 0.05), Negative (Score ≤ -0.05) or Neutral ($-0.05 \leq \text{Score} \leq 0.05$).

4) LSTM



In order to build the LSTM model for stock prediction,

- 1) Sequential for initializing the Neural Network.
- 2) Dense for adding a densely connected Neural Network layer
- 3) LSTM for adding the Long Short-Term Memory layer
- 4) Dropout for adding dropout layers that prevent Overfitting

We add the LSTM layer and later add a few Dropout layers to prevent overfitting.

We add the LSTM layer with the following arguments.

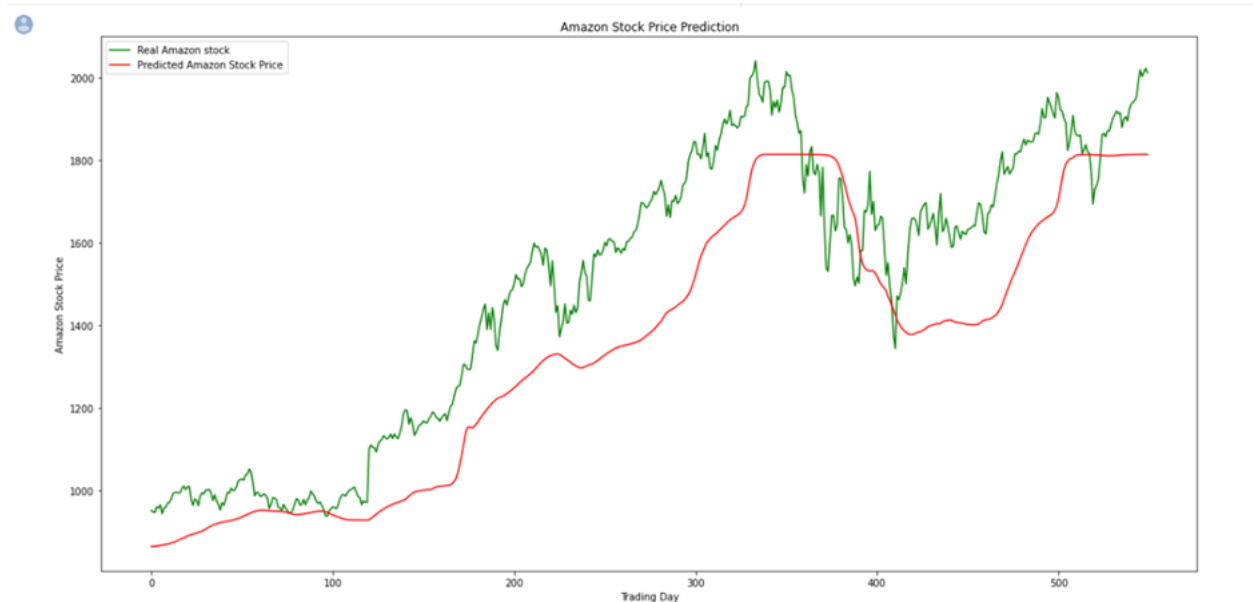
- 1) 64 units which is the dimensionality of the output space dimension
- 2) Input shape = (30,20) which is Number of Feature and Term Length.
- 3) Return_sequences = True which determines whether to return the last output in the output sequence, or the full sequence
- 4) In the Dropout layer, We specify 0.2 which means that 20% of the layers will be dropped to avoid Overfitting.
- 5) Batch normalization applies a transformation that maintains the mean output close to 0 and the output standard deviation close to 1. Which makes the network more stable during training.

Then, we apply a transformation that maintains the mean output close to 0 and the output standard deviation close to 1. Which makes the network more stable during training.

Results :

To know the significance of the External Data like News Sentiments, Trends data, Index Data (Dow Jones, S&P 500), We first trained our LSTM Model without using External Data features and Output obtained from that is shown below.

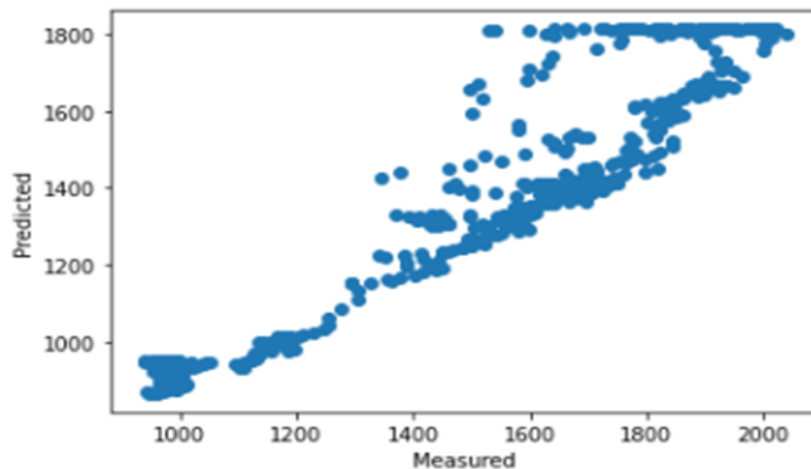
For Test Data :



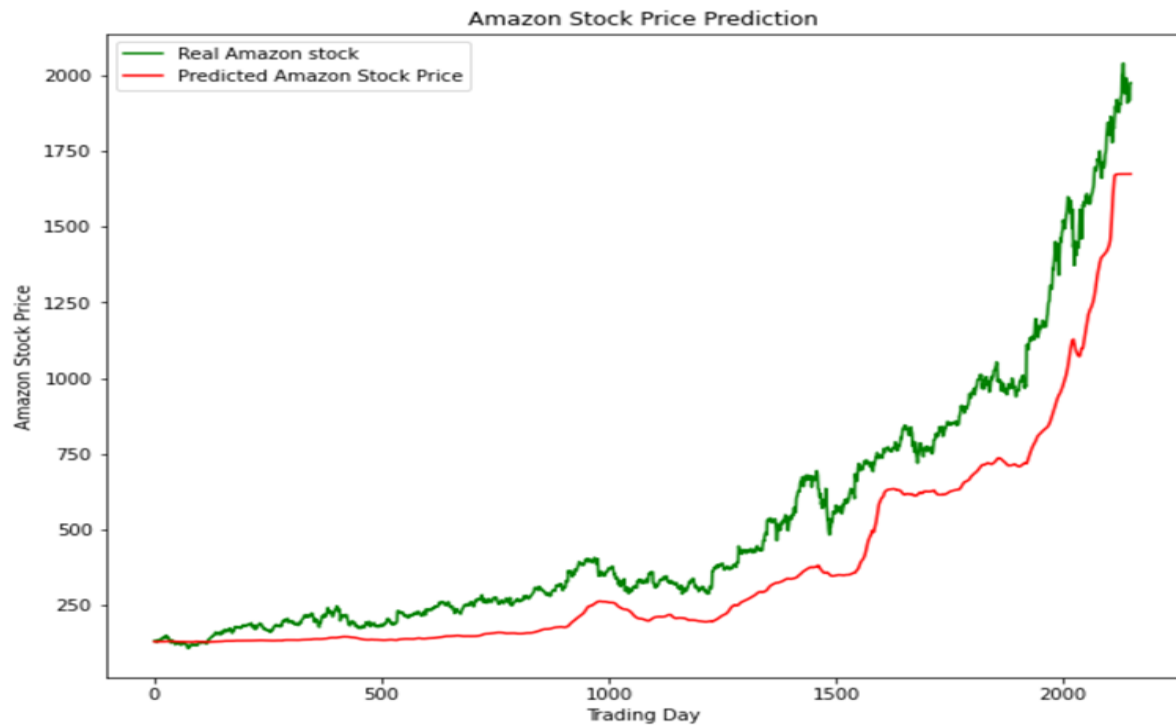
$$\text{MSE} = 34424.46366541709$$

$$\text{RMSE} = 185.53830781112856$$

$$\text{R2} = 0.7098161070736146$$



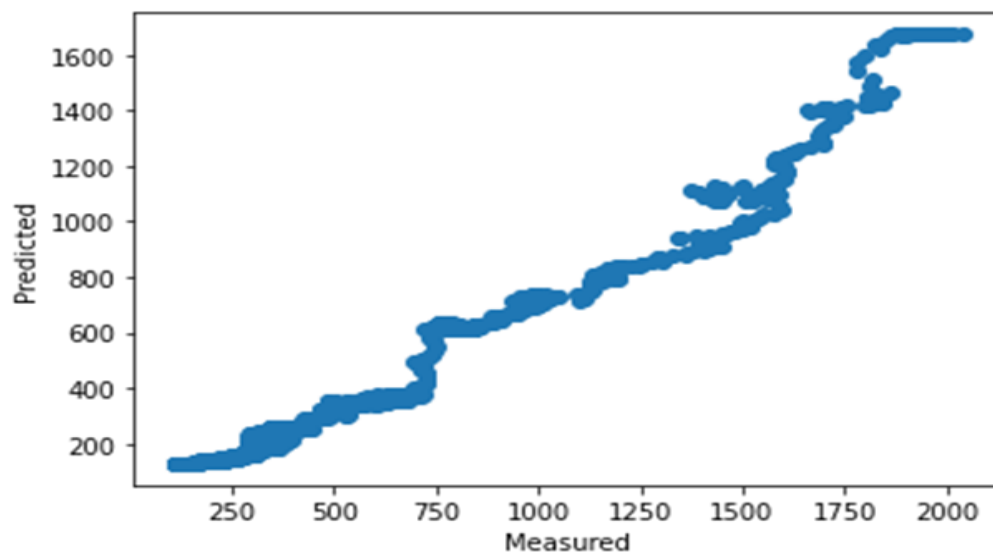
For Train Data :



MSE = 36206.13775168183

RMSE = 190.27910487408184

R2 = 0.8075211400734355



Now Considering External Data like News Sentiments, Trends data, Index Data (Dow Jones, S&P 500), We first trained our LSTM Model by also adding External Data features and Output obtained from that is shown below:

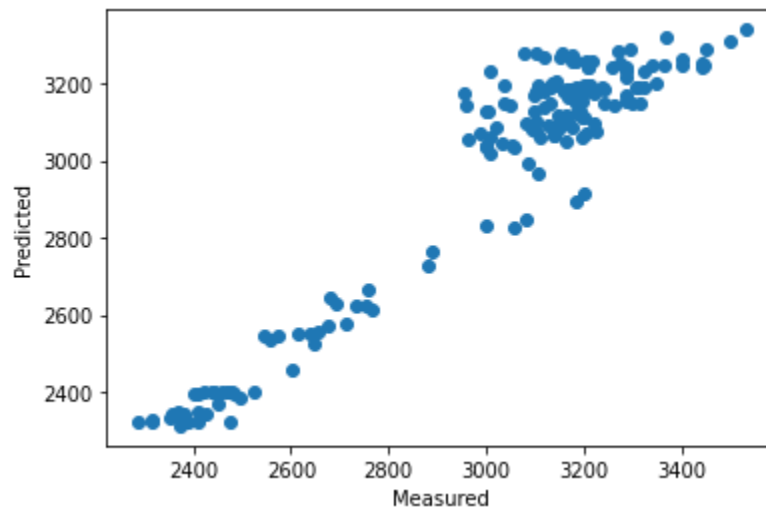
For Test Data :



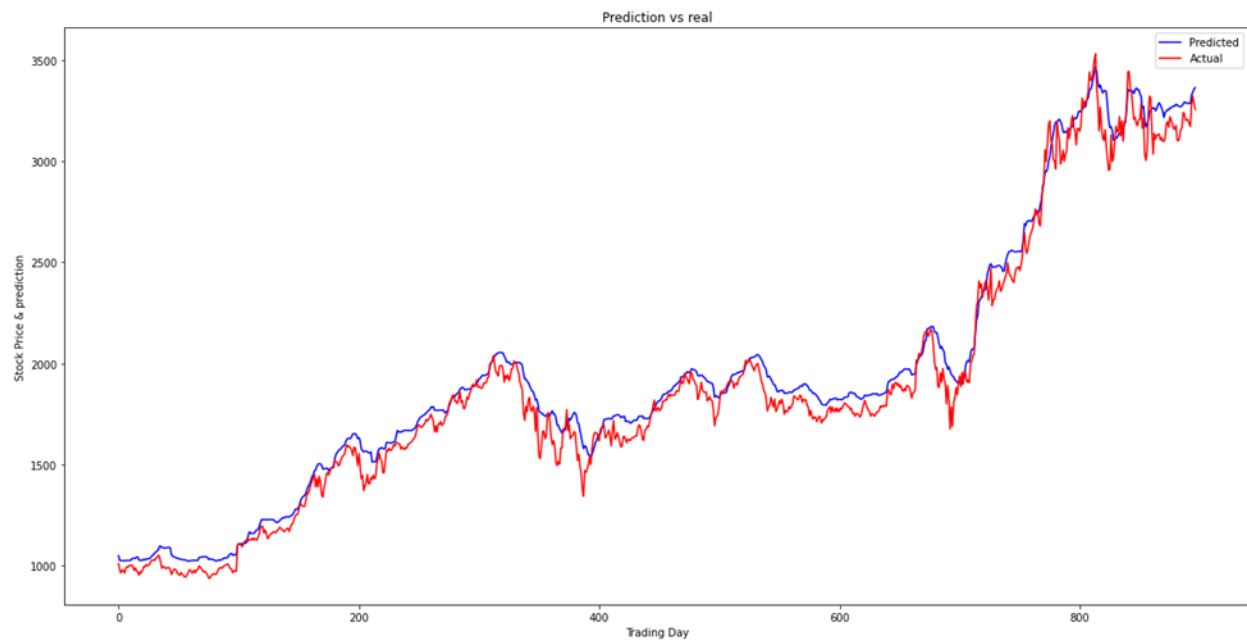
MSE = 10043.936767006682

RMSE = 100.21944305875324

R2 = 0.8997443916945977



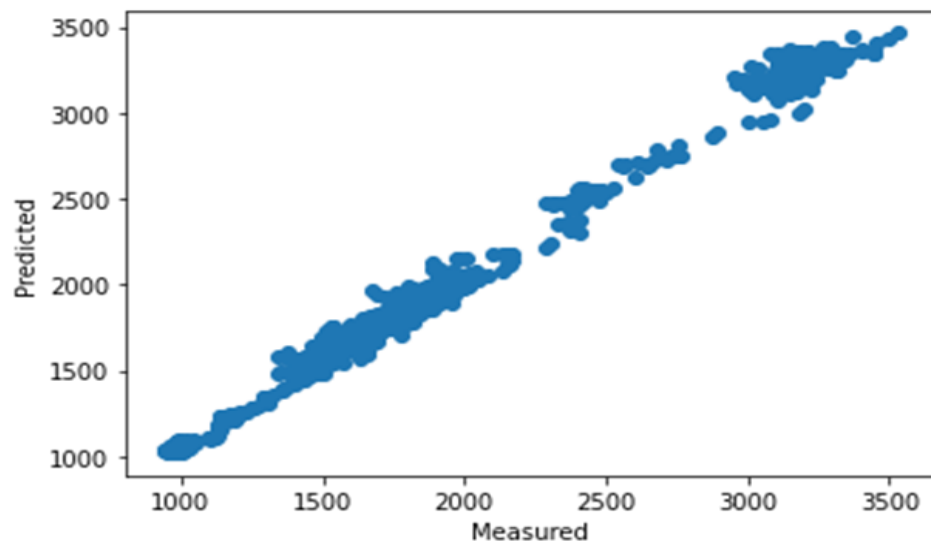
For Train Data :



MSE = 7329.326235687935

RMSE = 85.6114842511677

R2 = 0.9821823576560674



Conclusion :

From the above Result we can conclude that,

We can see the RMSE is Low and R2 Score is High, When we trained our Model also considering External Data like : Google Trends, Index Data and News Sentiment. Which shows their importance in the Prediction of the Stock Price.

Model's R2 score value is Low and RMSE is High for Train Dataset Compared with Test Dataset, Which is obvious and shows general behavior of ML Algorithms.

References :

- [Research Paper](#)
- [YFinance](#)
- [Macrotrends](#)
- [Seeking Alpha](#)
- [Google Trends](#)