STOCK MARKET ONLINE PRICE PREDICTION

A PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE AWARD OF THE DEGREE OF

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE ENGINEERING

SUBMITTED BY:

DEBAJYOTI RAY - 23043

SUMIT KUMAR SHARMA – 23921

SEMESTER- *4TH*

BRANCH- *CSE-2*

GUIDED BY

PROF. NAVEEN RATHI

SUBMITTED TO

DRONACHARYA COLLEGE OF ENGINEERING

FOR ACADEMIC YEAR 2022-21

CONTENTS

CHAPTER 1

INTRODUCTION

THE FINANCIAL MARKET IS A DYNAMIC AND COMPOSITE SYSTEM WHERE PEOPLE CAN BUY AND SELL CURRENCIES, STOCKS, EQUITIES AND DERIVATIVES OVER VIRTUAL PLATFORMS SUPPORTED BY BROKERS. THE STOCK MARKET ALLOWS INVESTORS TO OWN SHARES OF PUBLIC COMPANIES THROUGH TRADING EITHER BY EXCHANGE OR OVER THE COUNTER MARKETS.

THIS MARKET HAS GIVEN INVESTORS THE CHANCE OF GAINING MONEY AND HAVING A PROSPEROUS LIFE THROUGH INVESTING SMALL INITIAL AMOUNTS OF MONEY, LOW RISK COMPARED TO THE RISK OF OPENING NEW BUSINESS OR THE NEED OF HIGH SALARY CAREER. STOCK MARKETS ARE AFFECTED BY MANY FACTORS CAUSING THE UNCERTAINTY AND HIGH VOLATILITY IN THE MARKET.

ALTHOUGH HUMANS CAN TAKE ORDERS AND SUBMIT THEM TO THE MARKET, AUTOMATED TRADING SYSTEMS (ATS) THAT ARE OPERATED BY THE IMPLEMENTATION OF COMPUTER PROGRAMS CAN PERFORM BETTER AND WITH HIGHER MOMENTUM IN SUBMITTING ORDERS THAN ANY HUMAN. HOWEVER, TO EVALUATE AND CONTROL THE PERFORMANCE OF ATSS, THE IMPLEMENTATION OF RISK STRATEGIES AND SAFETY MEASURES APPLIED BASED ON HUMAN JUDGEMENTS ARE REQUIRED.

MANY FACTORS ARE INCORPORATED AND CONSIDERED WHEN DEVELOPING AN ATS, FOR INSTANCE, TRADING STRATEGY TO BE ADOPTED, COMPLEX MATHEMATICAL FUNCTIONS THAT REFLECT THE STATE OF A SPECIFIC STOCK, MACHINE LEARNING ALGORITHMS THAT ENABLE THE PREDICTION OF THE FUTURE STOCK VALUE, AND SPECIFIC NEWS RELATED TO THE STOCK BEING ANALYZED.

TIME-SERIES PREDICTION IS A COMMON TECHNIQUE WIDELY USED IN MANY REAL-WORLD APPLICATIONS SUCH AS WEATHER FORECASTING AND FINANCIAL MARKET PREDICTION. IT USES THE CONTINUOUS DATA IN A PERIOD OF TIME TO PREDICT THE RESULT IN THE NEXT TIME UNIT. MANY TIME-SERIES PREDICTION ALGORITHMS HAVE SHOWN THEIR EFFECTIVENESS IN PRACTICE.

THE MOST COMMON ALGORITHMS NOW ARE BASED ON RECURRENT NEURAL NETWORKS (RNN), AS WELL AS ITS SPECIAL TYPE - LONG-SHORT TERM MEMORY (LSTM) AND GATED RECURRENT UNIT (GRU). STOCK MARKET IS A TYPICAL AREA THAT PRESENTS TIME-SERIES DATA AND MANY RESEARCHERS STUDY ON IT AND PROPOSED VARIOUS MODELS. IN THIS PROJECT, LSTM MODEL IS USED TO PREDICT THE STOCK PRICE.

1.1 MOTIVATION FOR WORK

Businesses primarily run over customer's satisfaction, customer reviews about their products. Shifts in sentiment on social media have been shown to correlate with shifts in stock markets. Identifying customer grievances thereby resolving them leads to customer satisfaction as well as trustworthiness of an organization. Hence there is a necessity of an un biased automated system to classify customer reviews regarding any problem. In today's environment where we're justifiably suffering from data overload (although this does not mean better or deeper insights), companies might have mountains of customer feedback collected; but for mere humans, it's still impossible to analyze it manually without any sort of error or bias. Oftentimes, companies with the best intentions find themselves in an insights vacuum. You know you need insights to inform your decision making and you know that you're lacking them, but don't know how best to get them. Sentiment analysis provides some answers into what the most important issues are, from the perspective of customers, at least. Because sentiment analysis can be automated, decisions can be made based on a significant amount of data rather than plain intuition.

1.2 PROBLEM STATEMENT

Time Series forecasting & modelling plays an important role in data analysis. Time series analysis is a specialized branch of statistics used extensively in fields such as Econometrics & Operation Research. Time Series is being widely used in analytics & data science. Stock prices are volatile in nature and price depends on various factors. The main aim of this project is to predict stock prices using Long short term memory (LSTM).

CHAPTER 2

LITERATURE SURVEY

2.1 INTRODUCTION

"What other people think" has always been an important piece of information for most of us during the decision-making process. The Internet and the Web have now (among other things) made it possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics — that is, people we have never heard of. And conversely, more and more people are making their opinions available to strangers via the Internet. The interest that individual users show in online opinions about products and services, and the potential influence such opinions wield, is something that is driving force for this area of interest. And there are many challenges involved in this process which needs to be walked all over in order to attain proper outcomes out of them. In this survey we analyzed basic methodology that usually happens in this process and measures that are to be taken to overcome the challenges being faced.

2..2 The Stock Market and Investment

The research work done by Manh Ha Duong Boriss Siliverstovs. Investigating the relation between equity prices and aggregate investment in major European countries including France, Germany, Italy, the Netherlands and the United Kingdom. Increasing integration of European financial markets is likely to result in even stronger correlation between equity prices in different European countries. This process can also lead to convergence in economic development across European countries if developments in stock markets influence real economic components, such as investment and consumption. Indeed, our vector autoregressive models suggest that the positive correlation between changes equity prices and investment is, in general, significant. Hence, monetary authorities should monitor reactions of share prices to monetary policy and their effects on the business cycle.

CHAPTER 3

EXPERIMENT ANALYSIS

3.1 system configuration

This project can run on commodity hardware. We ran entire project on an Intel I5 processor with 8 GB Ram, 2 GB Nvidia Graphic Processor, It also has 2 cores which runs at 1.7 GHz, 2.1 GHz respectively. First part of the is training phase which takes 10-15 mins of time and the second part is testing part which only takes few seconds to make predictions and calculate accuracy.

3.1.1 Hardware Requirements:

- *RAM: 4 GB*
- *Storage: 500 GB*
- *CPU: 2 GHz or faster*
- *Architecture: 32-bit or 64-bit*

3.1.2 Software requirements

- *Python 3.5 in Jupyter Notebook is used for data pre-processing, model training and prediction.*
- *Operating System: windows 7 and above or Linux based OS or MAC OS.*

In [5]:
```
import math
import pandas_datareader as web
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from keras.models import Sequential
from keras.layers import Dense , LSTM
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
```

In [6]:
```
df = web.DataReader('AAPL' , data_source='yahoo', start='2010-8-22', end ='2021-8-22')
df
```

Out[6]:

	High	Low	Open	Close	Volume	Adj Close
Date						
2010-08-23	9.000000	8.758929	8.992500	8.778571	414041600.0	7.505835
2010-08-24	8.678571	8.523214	8.666786	8.568929	602565600.0	7.326587
2010-08-25	8.713929	8.471429	8.501429	8.674643	596867600.0	7.416974
2010-08-26	8.776786	8.581429	8.766071	8.581429	466505200.0	7.337275
2010-08-27	8.664643	8.412857	8.633929	8.629286	548391200.0	7.378193
...	...	...	...	...	...	...
2021-08-16	151.190002	146.470001	148.539993	151.119995	103296000.0	150.486649
2021-08-17	151.679993	149.089996	150.229996	150.190002	92229700.0	149.560562
2021-08-18	150.720001	146.149994	149.800003	146.360001	86326000.0	145.746597
2021-08-19	148.000000	144.500000	145.029999	146.699997	86960300.0	146.085175
2021-08-20	148.500000	146.779999	147.440002	148.190002	60549600.0	147.568939

In [8]:
```
plt.figure(figsize=(16,8))
plt.title('Close Price History')
plt.plot(df['Close'])
plt.xlabel('DATE', fontsize=18)
plt.ylabel('Close Price USD ($)',fontsize=18)
plt.show()
```

In [11]:
```
scaler=MinMaxScaler(feature_range=(0,1))
scaled_data=scaler.fit_transform(dataset)
scaled_data
```

Out[11]:
```
array([[1.47064779e-03],
       [0.00000000e+00],
       [7.41585784e-04],
       ...,
       [9.66608496e-01],
       [9.68993580e-01],
       [9.79446014e-01]])
```

In [12]:
```
train_data=scaled_data[0:training_data_len,:]
x_train=[]
y_train=[]
for i in range(60,len(train_data)):
    x_train.append(train_data[i-60:i,0])
    y_train.append(train_data[i,0])
    if i<= 60:
        print(x_train)
        print(y-train)
        print()
```

```
[array([1.47064779e-03, 0.00000000e+00, 7.41585784e-04, 8.76865363e-05,
       4.23406821e-04, 6.43877547e-04, 7.94203058e-04, 2.60558198e-03,
       3.06656683e-03, 4.72011400e-03, 4.47959987e-03, 5.75984204e-03,
       5.79742007e-03, 5.88260453e-03, 6.79205045e-03, 7.04759713e-03,
       7.58875559e-03, 9.17966379e-03, 8.87902614e-03, 1.08482435e-02,
       1.09835298e-02, 1.19806689e-02, 1.22737936e-02, 1.31256181e-02,
       1.28349955e-02, 1.17576894e-02, 1.18854628e-02, 1.09785256e-02,
       1.06703617e-02, 9.69828345e-03, 1.22788112e-02, 1.23414368e-02,
       1.23489564e-02, 1.35640575e-02, 1.38872473e-02, 1.46839591e-02,
       1.50848205e-02, 1.56284877e-02, 1.87426543e-02, 1.95594095e-02,
       1.74273362e-02, 1.76878937e-02, 1.74348558e-02, 1.69212537e-02,
       1.72644869e-02, 1.70665683e-02, 1.70114490e-02, 1.63625606e-02,
       1.52952695e-02, 1.60969855e-02, 1.73947623e-02, 1.82566152e-02,
       1.96270526e-02, 1.93414342e-02, 1.97147392e-02, 1.90783746e-02,
```

In [19]:
```
model.compile(optimizer='adam',loss = 'mean_squared_error')
```

In [20]:
```
model.fit(x_train,y_train,batch_size=1,epochs=1)
```

```
2156/2156 [==============================] - 25s 11ms/step - loss: 1.9481e-04
```

Out[20]:
```
<keras.callbacks.History at 0x273ffabc8b0>
```

In [21]:
```
test_data=scaled_data[training_data_len-60:,:]
x_test=[]
y_test=dataset[training_data_len:,:]
for i in range(60 , len(test_data)):
    x_test.append(test_data[i-60:i,0])
```

In [22]:
```
x_test=np.array(x_test)
```

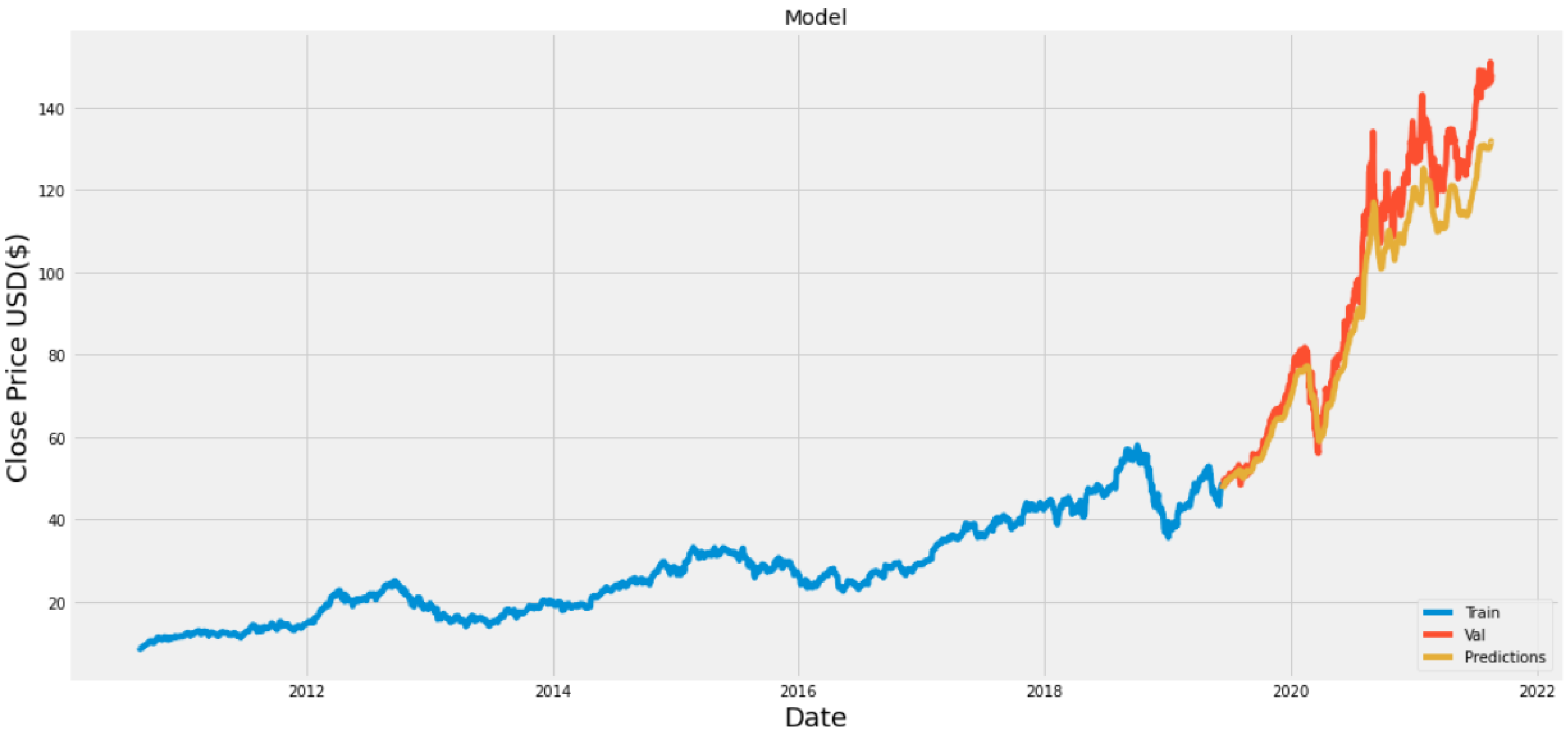
In [23]:
```
x_test=np.reshape(x_test,(x_test.shape[0],x_test.shape[1],1))
```

In [26]:
```
predictions=model.predict(x_test)
predictions=scaler.inverse_transform(predictions)
```

```
18/18 [==============================] - 0s 8ms/step
```

In [27]:
```
rmse=np.sqrt(np.mean(((predictions-y_test)**2)))
rmse
```

Out[27]:
```
9.37113640232443
```



Model
Close Price USD($)
140
120
100
80
60
40
20
2012
2014
2016
2018
2020
2022
Date
Train
Val
Predictions

In [31]:

```
apple_quote= web.DataReader('AAPL',data_source='yahoo',start='2010-08-15',end='2021-08-15')
new_df=apple_quote.filter(['Close'])
last_60_days=new_df[-60:].values
last_60_days_scaled=scaler.transform(last_60_days)
X_test =[]
X_test.append(last_60_days_scaled)
X_test=np.array(X_test)
X_test=np.reshape(X_test,(X_test.shape[0],X_test.shape[1],1))
pred_price=model.predict(X_test)
pred_price=scaler.inverse_transform(pred_price)
print(pred_price)
```

```
1/1 [==============================] - 0s 15ms/step
[[130.73286]]
```

In [32]:

```
apple_quote2= web.DataReader('AAPL',data_source='yahoo',start='2021-08-16',end='2021-08-16')
print(apple_quote2['Close'])
```

```
Date
2021-08-16    151.119995
Name: Close, dtype: float64
```

CHAPTER 4

CONCLUSION AND FUTURE WORK

4.1 Conclusion

In this project, we are predicting closing stock price of any given organization, we developed a web application for predicting close stock price using LMS and LSTM algorithms for prediction. We have applied datasets belonging to Google, Nifty50, TCS, Infosys and Reliance Stocks and achieved above 95% accuracy for these datasets.

4.2 Future work

- *We want to extend this application for predicting cryptocurrency trading.*
- *We want to add sentiment analysis for better analysis.*