STOCK MARKET PREDICTION USING MACHINE LEARNING AND PYTHON

A Project

by

MANOJ M, LAKSHIN D K, MUYEen baig,

manohar s n, nagsharan, akhilesh, mohan parakash v

2ND YEAR

INFORMATION SCIENCE AND ENGINEERING

VIDYAVARDHAKA COLLEGE OF ENGINEERING

Submitted to the CSRBOX team

as the final project deliverable for

IBM SkillsBuild Data Analytics Innovation Camp

**Project Summary**

It has never been easy to invest in a set of assets, the abnormally of financial market does not allow simple models to predict future asset values with higher accuracy. Machine learning, which consist of making computers perform tasks that normally requiring human intelligence is currently the dominant trend in scientific research. This Project aims to build a model using Recurrent Neural Networks (RNN) and especially Long-Short Term Memory model (LSTM) to predict future stock market values. The main objective of this project is to see in which precision a Machine learning algorithm can predict and how much the epochs can improve our model.

Several studies have been the subject of using machine learning in the quantitative financial, predicting prices of managing and constricting entire portfolio of assets, as well as, investment process, and many other operations can be covered by machine learning algorithms. In general machine learning is a term used for all algorithm’s methods using computers to reveal patterns based only on data and not using any programming instructions. For quantitative finance and specially assets selections several models supply a large number of methods that can be used with machine learning to forecast future assets value. This type of models offers a mechanism that combine weak sources of information and make it a strange tool that can be used efficiently. Recently, the combination of statistics and learning models have polished several machine learning algorithms, such as acritical neural networks, gradient boosted regression trees, support vector machines and, random forecast. These algorithms can reveal complex patterns characterized by non-linearity as well as some relations that are difficult to detect with linear algorithms. These algorithms also prove more effectiveness and multi collinearity than the linear regressions ones.

A large number of studies is currently active on the subject of machine learning methods used in finance, some studies used tree-based models to predict portfolio returns, others used deep learning in the production of future values of financial assets. Also, some authors overviewed the forecasting of returns using of other algorithm. Others proceeds to forecast stock returns using unique decision-making model for day trading investments on the stock market the model developed by the authors use the support vector machine (SVM) method, and the mean variance (MV) method for portfolio selection.

In our project we are using other Machine Learning approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. Specifically, we are using Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies. In the whole project we are using only Line graph for visualization.

# TABLE OF CONTENTS

CHAPTER

1. INTRODUCTION 4

Topic Brief 4

Purpose of the Study 4

Hypotheses 5

Significance of the Study 5

Method of Procedure 6

Collection of Data 6

Treatment of the Data 6

Data Source 6

Definitions of Terms 6

2. PRESENTATION OF FINDINGS (or DATA) 7

3. SUMMARY OF THE STUDY AND THE FINDINGS, CONCLUSIONS, ESTIMATIONS, PREDICTIONS, FUTURE COURSE OF ACTION 10

REFERENCES 10

APPENDICES 11

Appendix

Chapter 1

INTRODUCTION

**Topic Brief**

The Stock market check is an exceptionally fascinating errand which joins high substances of how the budgetary exchange limits, and what unconventionalities can be prompted in a market in light of different conditions. While a few venders may battle that the market itself is functional, and that if there is new check or any assortment from the standard in a market it charms it by auditing itself, thusly making no space for conjectures, while several vendors may battle that on the off chance that the information is orchestrated well, by then machine can make a sort out of procedure that is persuading can affect high continue exchanging or HFT, which is just conceivable through Algorithmic Trading Systems or Automated Systems of Trade.

Money related authorities think about the expression, buy low, move high yet this does not give enough setting to settle on proper endeavor decisions. Before an investigator places assets into any stock, He should realize how money markets continues. Setting assets into a wonderful stock regardless at a horrible time can have awful results, while vitality for a common stock at the fortunate time can hold up under focal points. Cash related monetary pros of today are going toward this issue of trading as they don't for the most part understand concerning which stocks to buy or which stocks to offer with the authentic objective to get impeccable focal points. Envisioning whole game plan estimation of the stock is commonly clear than foreseeing on day-to-day premise as the stocks change rapidly reliably subject to world events.

The answer for this issue requests the utilization of instruments and advances identified with the field of information mining, design acknowledgment, machine learning and information forecast. The application will foresee the stock costs for the following exchanging day. The necessities and the usefulness of this application corresponds it to the class

**Purpose of the Study**

Artificial intelligence (AI) to play an integral role in our day-to-day life applications whether it be home environment applications like Alexa or financial applications like trading, it is a development towards a new era of technology. This project comprises of an application of AI on financial data, known as algorithmic trading.

Automated trading systems involves the use of complex AI systems to make extremely fast trading decisions like buy, hold, or sell. It involves high frequency trading or HFT to make millions of trades in a day.

Machine learning is a subset of AI and generally provides solutions which learn from experience without being explicitly programmed.

In simple words, just the machine learning models are selected and fed with data the model then automatically adjusts its parameters and improves its outcome.

**Hypotheses**

* Stock market prediction is the act of trying to determine the future value of the company.
* The programming language is used to predict the stock market using machine learning is Python.
* we propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction.
* Recurrence Neural Network (RNN) to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies.
* Finally, We Use the data Visualization technique to compare the Actual and the predicted stock prices.

**Significance of the Study**

1. Analyze a dataset:

We are using the Google Stock price which is the historical data that we are using entire project. We load the data to our file using pandas library of Python. First, we need to clean the data that is we need to see that the data contains Null or Nan values. We checked that there is no any null value. The Data contains 1258 rows and 5 columns the name of the columns is Date, Open, High, Low, Close, Volume. Then we check the data type of all columns to maintain the Homogeneity we convert all the columns to the same data types i.e., float. By this we are done with Data pre-processing, Data Normalization and data regularization steps.

1. Prediction and estimation.

Now we start visualizing the data first we see the growth of the data by considering the open column in the we sow that there is growth of 85% of price from the year of 2012 to 2017. Next, we find 7 days rolling mean by pandas which means for every stock prediction we look 7 days back and get the average of our column and we compare that by visualization the we build RNN and We trained it and get the predicted stock prices.

1. How can this analysis bring a significant value?

Once we done with the training, we start testing the model by test data then we put into our model and get the output. At last, we visualize real and predicted data to check weather the prediction is correct are wrong.

**Method of Procedure:**

**Collection of Data**

We are developing model to predict the future stock prices hence we need the recorded stock data. We got the Google stock data from the website called Kaggle. This Data is ranges form the year of 2012 to 2017, this data is used for training purpose. For testing we had used only 2017 dataset.

**Treatment of Data**

Once we got the Data, we need to check whether it is good or not. Which means we need to check it contains Null or Nan values. We checked it but it does not contain any missing values. Once it is done, we need to convert all the columns to the same data type to maintain the homogeneity. Then we regularize the data.

**Definitions of Terms**

Columns that are present in the data are Data, Open, High, Close, Low, Volume. Description of the data is given below

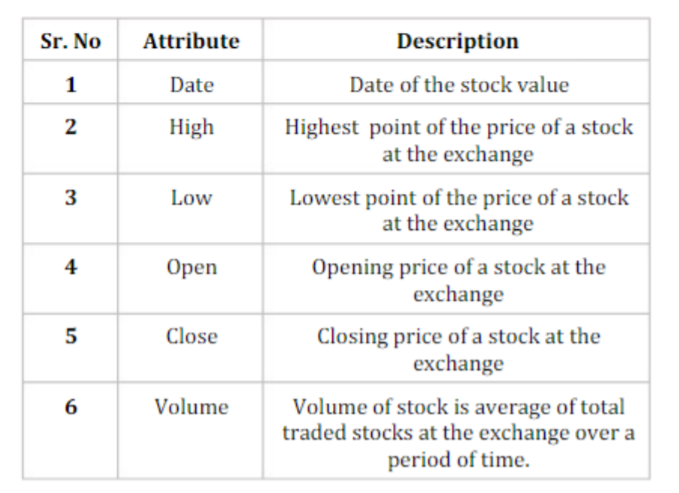


Fig 1: Data column description

Chapter 2

**PRESENTATION OF FINDINGS**

Initially we started with cleaning the data which is the initial step of the data preprocessing. Which means we checked that is there any missing values or any Nan values by some pandas and NumPy functions. Ones the cleaning is done we move to the data normalization technique that is we get the datatype of the every column and we convert all column to the same datatype in this case that is float .

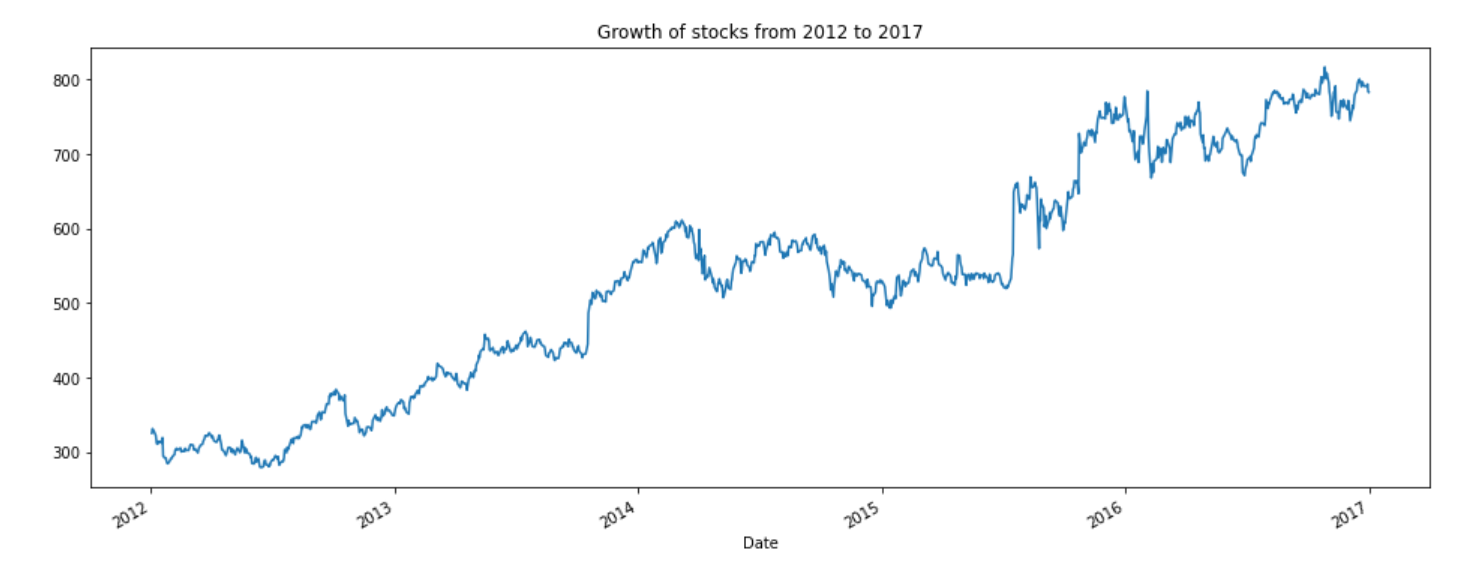
Once the Data cleaning is done, we plot the line graph for the column Open to see the growth of the stock. The below fig (fig-2) shows the growth of the stock.

Fig 2: Growth of stock from 2012 to 2017

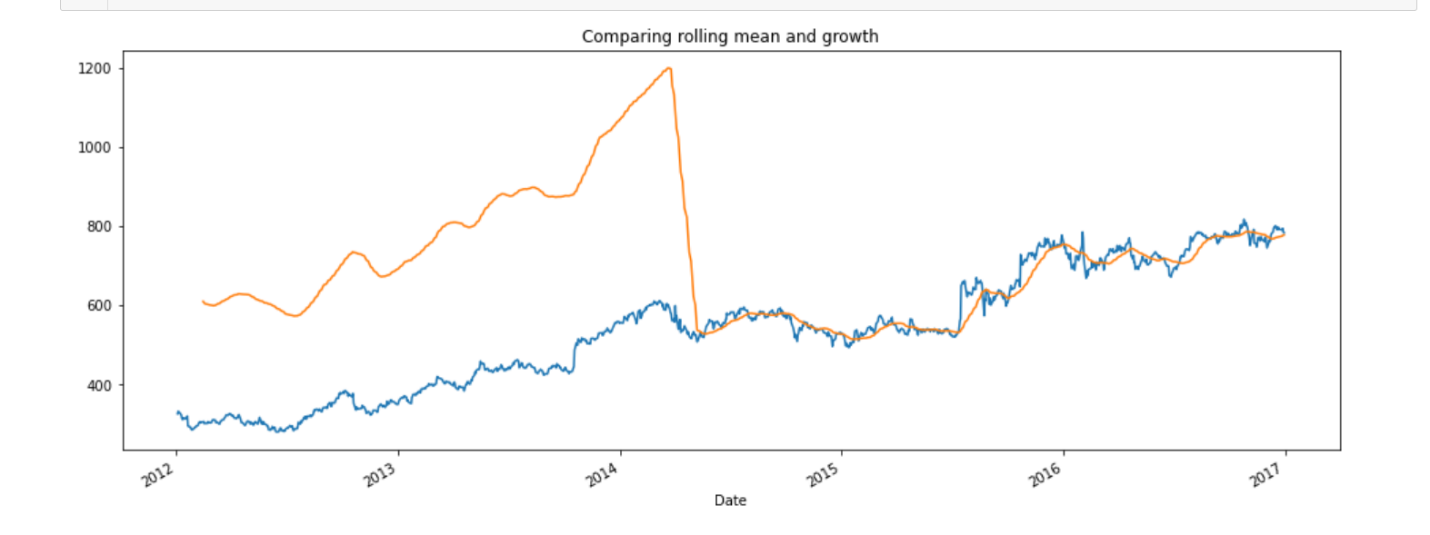
 From the above graph we can see the there is an 85% increase of stock from 2012 to 2017. After this we calculated rolling mean of the 7 days which means in the simple example in every single stock prediction look 7 days back collect all the transactions that fall in this column and get the average of the 7 days.

Fig 3: Comparing Rolling mean and growth

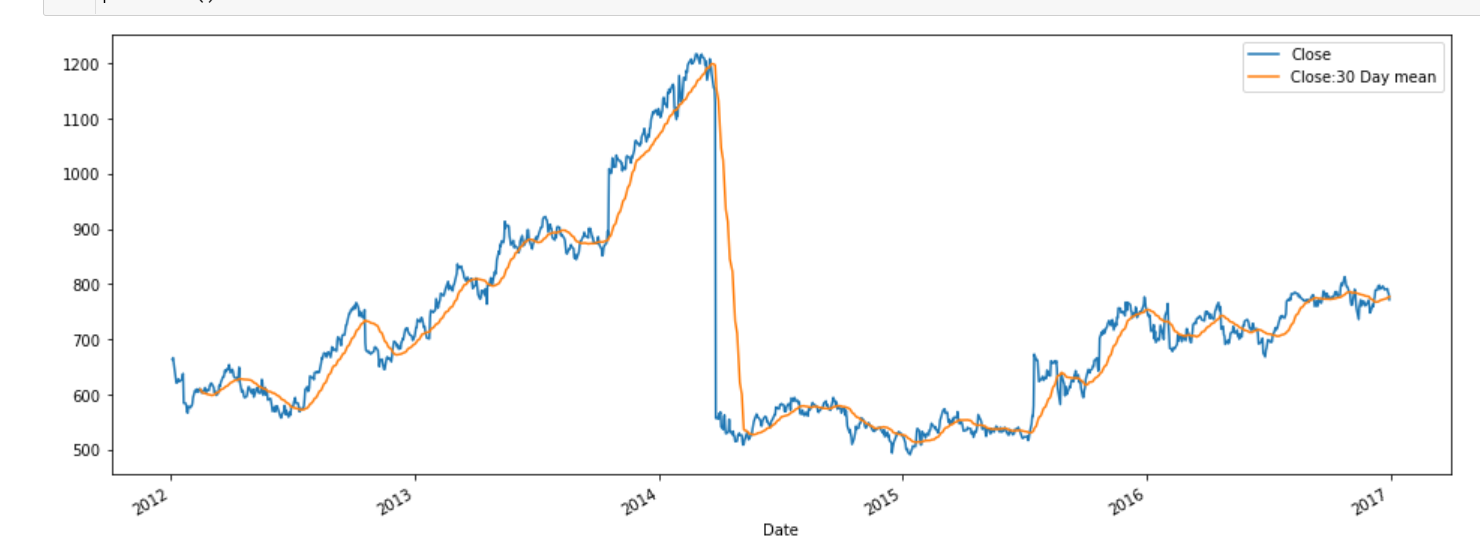
 Then we plot the line graph between Rolling mean and the growth of the stock. The above fig (fig 3) shows that graph. Blue indicates the growth and orange indicates rolling mean Then we di same process to the Close column the visualization is shown below

Fig 4: 30 days rolling mean for Close column

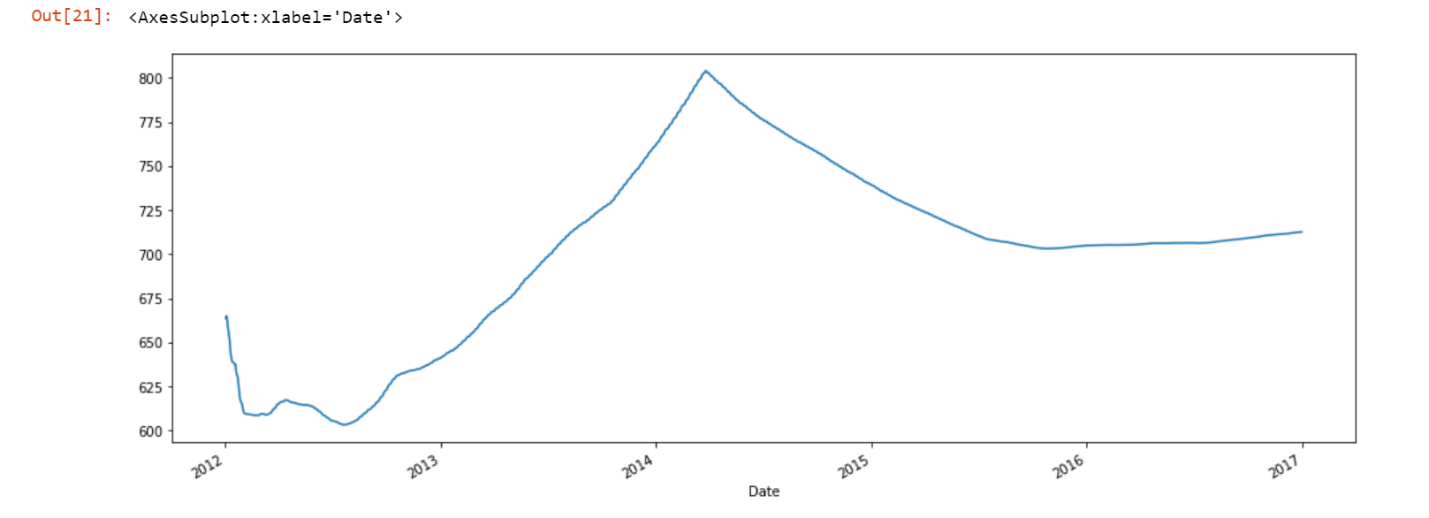
 Next, we plot the line plot for min period is equal to 1. Basically, we are going to say that the minimum number of observations per window which is of 30 days is 1 and the graph is shown below at fig 5.

Fig 5: When min period is 1

Now we build the first Data Frame that is the training dataset and reading the contents of data set using pandas And reading the contents of the data set using Pandas to do when we take into account in everything from the start of the time series to the golden point of value now we are going to be data preprocessing we are will be using the min Max scaler to transform features bifurcating each of them to set range and here featuring is zero to one for these reasons than we can go ahead and when it then finally we are going to be creating a data structure with 60 the day and one output to basically what we're trying to do sure is that it is really going to take the data from the wanted a 60 and then make predictions of the 61st day.

Then we are going to for the top by taking data from a number to today Number 61 and protect on the sixties s t o r y is going to go from 60 to the end of a range was 1258 and then we are going to appended to the extreme why train extreme starts at a - 60 days starts from 1 -6 and that 61 and the Y train is basically when to give us our prediction on the identity which if we take 61 is a first prediction from the for 60 days so while train is going to reverse the protection of the 61st day the preprocessing data discretization where will reduce upon 240 top 8 with particular importance for the numerical data next able to do it up transformation which is basically nothing but an organization that will clean the villain the missing values into it audit of files and after the data set is transformed into clean data set the data will be divided into training and testing steps was to evaluate so in the end what we are trying to do here is creating a data structure with 60th day and one output in total start out by training at data doing the same thing which we have done before checking if there is any not appear to be possibilities and then move on to feature scaling for which we are going to be important minmaxscaler from scikit learn which is nothing but a machine learning library for python.

Start by importing heroes libraries and packages this is going to be the first step towards building your own and is basically tenth of close high-level API libraries first is sequential basically a linear stack of layers through which you can create a sequential model by parcel destroyed when we are going to import something known as a sentence is the regularly connected neural network layer is the most frequently used to live and it used to change the dimensions of you’re a goodbye the dental basically represents a matrix vector multiplication the values in the crisis which country Nobel parameters get updated during backpropagation so if you get an Empty dimensional vector as output a dense layer is used to change the dimensions of your mind that they want to be initialized in our own and for a time series problem like this we are basically going to be using the regression model for regression deep learning model the first step is to read in the data which is a sequential data and there were assigned to the model called impressive with that we are moving on to stage for which is the most important area that is trained in the neural network in the stage your data is going to be fed to the neural network model for prediction you can assign random pic is and base to your model not the LSTM what about basically is a regularization technique for reducing overfitting in neural networks so it drops out units in a neural network then you have you three layers model is composed of a sequential input layer followed by three layers and enslaved with activation and then finally a dense output layer with the linear activation function so here you can see the first input layer you have your units return sequences the input shape and then the dropout yes and then finally you have your output here and since you need only one output unit is canvas model another type of Optimizer use can greatly affect how fast the algorithm converges to the minimum value also it is important 1 next what you going to do is compile your own and hear you can be using something known as an Optimizer and let me just begin by saying that and Optimizer is one of the two arguments that are required for reaching the Global minimum number of you played algorithms that I've chosen to use the Adam Optimizer which combine The Perks of the ADAgrad And The RMS prop.

Finally, there were no use metric left to visualize the result of the product in stock and the real stock price the plot you can see that the real stock price when top model was predicted that the price of the stock will work this clearly shows how powerful LSTM are. us for analyzing the time series and sequential data for the analysis you have been implemented with relative ease thanks to Cairns and the functional API.

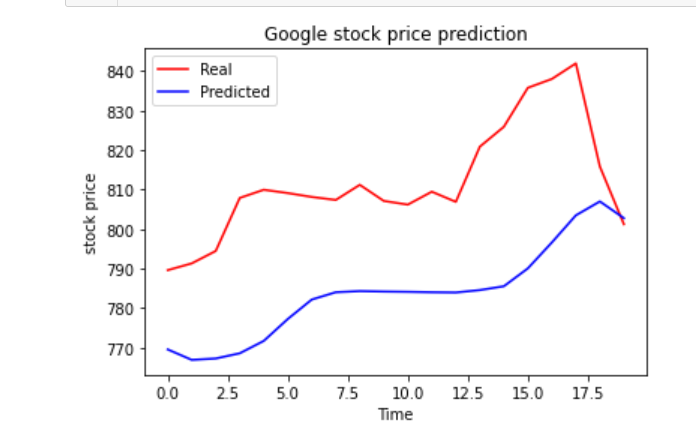
The below fig (Fig 6) Shows the Final visualization.

Fig 6: Real Stock price v/s Predicted Stock price

**Conclusion**

conclusion I like to say that the popularity of stock market trading is going extremely rapidly which is encouraging researchers to find out new methods for the prediction using Newton name the forecasting technique is not only helpful to the soldiers on to also helps investors or any person dealing with the stock market in order to help predict stock indices of forecasting model with the good accuracy is required in this work we have used one of the most precise forgotten Technologies using RNN and LSTM unit which have investors analyst or any person interested in investing in stock market by providing them a good knowledge of the future situation of the stock market time at least we were able to get the trend .

REFERENCES

**Data set link:**

<https://www.kaggle.com/medharawat/google-stock-price>

APPENDICES

Code of the project:

