STOCK PRICE PREDICTION USING DEEP LEARNING AND TIME SERIES MODELS

STOCK PRICE PREDICTION USING TIME SERIES MODELS

Dissertation submitted in partial fulfillment of the requirements

for the degree of

MSc Data Analytics

At Dublin Business School

Homanshu Dalal

Student Number - 10394591

Supervisor: Mr. Shahram Azizi Sazi

DECLARATION

I, Homanshu Dalal, declare that this research is my original work and that it has never been presented to any institution or university for the award of Degree or Diploma. In addition, I have referenced correctly all literature and sources used in this work and this work is fully compliant with the Dublin Business School's academic honesty policy.

ACKNOWLEDGMENTS

I would like to express my special thanks of gratitude to Mr. Shahram Azizi Sazi my teacher and research supervisor for his patient guidance from the beginning with constructive and valuable suggestions provided throughout the planning and improvement of my dissertation. She also helped me in doing a lot of Research, and her ability to give his time generously has been particularly valued.

Secondly, I would like to thank my family and friends for their support and encouragement throughout my course

Table of Contents

List of Figure

Abstract

Stock market price prediction is one of the complex difficult Task to predict. Because it is not only depending on one factor which gives you the correct prediction. So many researches have been done research on this but till now no single method has been invented which resolve this issue. This thesis presents the comparison between the four algorithms which clearly distinguish the which algorithm works better for google stock market prediction. In this paper, four algorithms have been used which are LSTM, ARIMA, AR and Reinforcement learning. Which clearly gives us the idea that Time series algorithms works better than LSTM algorithm. This paper will help to many trading investors to invest their money with proper decision so they can make the profits. Main reason behind using these algorithms are stock market dataset is relating to the time series data where these algorithms work effectively.

Chapter 1: Introduction

1.1 Background of Dissertation Topic

Stock price is for single stock which are sold by different public offering companies. Having stocks of any company allows you to own portion of it. Usually owner of company sells their stocks to get additional money to invest in their business to expand their business. Initial publishing the stocks at stocks market is called Initial Public Offering (IPO).

Stock price is not the same all time. it varies according to supply and demand. If number of people is capable to buy stock, then chances of price going high is very high. If so many people ready to sell bought stocks, then price of stocks goes down as there is more supply than demand. It's very clear now the supply and demand and relatively proportional to each other. Not only this factor is responsible for up and down stocks price but also economic factors like market behavior, trends and specially image of company, which is directly depend on the news channel. What kind of news being showed whether is it positive or negative.

Another motivation of this research is that it contains numerous hypothetical and test challenges. Efficient Market Hypothesis is one the main crucial hypothesis [1]. Many machine learning models have built by so many scientists for stock market predictions, using different types of fundamental technical and analytical methods. Analytical method is used to find out the many small economic factors which address to the effect of stock prices. On other side technical analyst fixes the utilizing value, volume and other financial factual graphs to foresee stock developments. The calculated affected markets cost which influence by external and internal part of great technical analysis [2].

Behind all the approaches made by scientist the common approach is machine leaning algorithms. There lots of algorithms has been used to predict the prices but here in this research, four algorithms have used which are RNN-LSTM, TIME SERIES – ARIMA, AR and Reinforcement learning. RNN has better performance in time based sequential data but it has gradient decent problem of exploding and vanishing. Thus, it does not capture long term dependencies. But recently LSTM architecture has showed excellent performance to overcome this limitation [5].

1.2 Investment Theory

If we want to know how the stocks is formed, how the stock behaves and what is the shape? Then two theories are very essential first one is, Firm Foundation and second is Castles in the Air [4]. This theory not only limited to these questions but also tells us what parameter needs to be taken seriously before investing his money into market. It very crucial to know this fact because, without knowing all these factors if someone invest money into market, he might have face big loss. Hence these theories are very important. There is not doubt these theories tells to think and react. This sequence might give idea to investors how can they allocate the money in the capital stock market. Every people who invest their money, they want profit. And no doubt profit is dependent on investing in stocks which have good future. Good future means the companies who have very good performance since long past time. And they will continue

their performance in future excellent. As the company performs well surely their price of stock also goes high and gives good profit. But sometimes people need to think that what are they trying to achieve one way or other is to predict the future of market. But what is that factors who determines the future? The only information that people have. Hence, we have the following schema in Figure 1.

Schema defines few points which are very important to understand until what time we need to hold the share. When have we to take decisions about buy and sell? Investors "reacts" when he has the knowledge of component "Information" according to the Firm Foundation theory market is made from the reaction of investors and it is poked by the real value of firms. But how real value is calculated what is that? It is very careful analysis about the present condition of organization and future goals [3]. That defines the real value. It means if one company's share continuously going to down and without study someone buys it, surely tomorrow's share price will lower than today's hence its very important find out the real values of companies. On other hand Castles in Air suggest investors to buy stock by investor's behavior. This theory doesn't care about the future goals of company. Investor is should have to buy the stock with minimum stock price and sell tomorrow if the stock price goes high. Investors does not need to think about the further future price of stocks. Hence, we can say that Firm Foundation theory is defined by the mostly logic and other Castles in the Air theory is defined by the mostly psychology

1.3 Problems Defining

Based on the historical data some forecasting can be done for some future events is called forecasting. It includes many areas such as industry and business, economics, environmental-science and finance. And their problems can be classified as

- Short term forecasting (prediction for few seconds, minutes, days, weeks or months)
- Medium term forecasting (prediction for 1 to 2 years)
- Long term forecasting (prediction beyond 2 years) [6]

Time is the main feature in while predicting stock price. It needs to be very carefully analyzed. Time series data is usually defined as chronological sequence of observations for selected variable. In this research variable is stock price it can be univariate or multivariate. When information is about only one stock then is called univariate and data includes the stock price more than one organization the is called multivariate for multiple instances of time. what make trader understand where to invest their money which is analysis of time series data which helps to understand the trend, identifying patterns, trends and periods or cycles in the existing data. An early knowledge of bullish or bearish help investor to use their money very wisely not only this but also helps to find out the which is the best company where investor can invest their money for specified period. Thus, prediction and analysis of time series makes an important area of research. [7]

The Existing Method for stock price prediction can be classified as follows

- Fundamental Analysis
- Technical Analysis
- Time Series Forecasting

The value of company is estimated by fundamental analysis of investment of share analyzed by sales, profits, earnings and economic factors. This is one of the methods which suits to the predict the long-term forecasting. Technical analysis uses the historical price of stocks for identifying the future price. For technical analysis moving average algorithm commonly used. Unweighted mean of past n data points is suitable to short term predictions. And last one method is for analysis of time series data. It has two classes of algorithms they are

1. Linear Models
2. Non-Linear Models

The various types of linear models are AR, ARMA, ARIMA and its different variations [8] [9] [10]. To fit mathematical model these model uses some predefined equations to univariate time series. The interdependencies of other stock are not identified by these algorithms model since they deal with the only univariate time series. Also trained model for one time series will not compulsory fit for other time series data. Due to this reason it not possible to identify the trending patterns.

In the case of stock market, generated data is highly enormous and no linear. And to trained and model such kind of data we need robust model which can learn the hidden patterns and underlying dynamics. In such scenario deep learning algorithm model can effectively train and idenfying and exploiting the interactions and patterns existing in a data through a self-learning process. Unlike other algorithms deep learning algorithm works very well and provides good prediction by analyzing the hidden patterns within the data. In [11], there are so many applications of deep learning models are available for multivariate time series analysis. The first attempt has made to model for financial time series analysis using neural network was introduced in [12]. The goal behind making this model is decoding the non-linear regularities in asset price movements for IBM. Even if scope of this model is limited but it helped in establishing evidences against EMH[13]

Chapter 2: Literature Review

2.1 Business Importance

Stock market provides two important purposes. The first and important is to provide capital to business so they can use those funds and can grow their business. If company issues 2 million of stock in stock market with initial each cost of stock is 20$, so ultimately, they are getting $20 million money to invest into their business to expand it very rapidly. It is very feasible for company because if company borrows money from banks, they have pay back interest occurred on the capital amount. Company can avoid all these interest by using the amount which they get from share market.

The secondary purpose the stock market is for investors – those investors who purchase stocks they get opportunity to take the share profit from publicly traded companies. Usually investors get the advantage when dividends happen regularly on one stock. Secondly investors can sell their stock when bought stock goes up than purchased price, so they get the profit from them.

2.2 Objective

As shown in figure 1, Schema defines few points which are very important to understand until what share need to be hold. When to take decisions about buy and sell? Investors "reacts" when he has the knowledge of component "Information" according to the Firm Foundation theory market is made from the reaction of investors and it is poked by the real value of firms. But how real value is calculated what is that? It is very careful analysis about the present condition of organization and future goals [15]. That defines the real value. It means if one company's share continuously going to down and without study someone buys it, surely tomorrow's share price will lower than today's hence it's very important find out the real values of companies. On other hand Castles in Air suggest investors to buy stock by investor's behavior. This theory doesn't care about the future

goals of company. Investor is should have to buy the stock with minimum stock price and sell tomorrow if the stock price goes high. Investors does not need to think about the further future price of stocks. Hence, we can say that Firm Foundation theory is defined by the mostly logic and other Castles in the Air theory is defined by the mostly psychology.

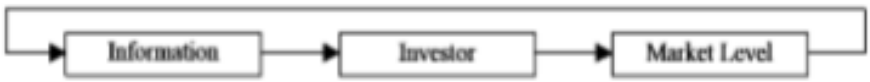


Figure 2.2

2.3 Previous Research Analysis

The lots of work has been done on the stock market prediction but till now neither enough nor dependent algorithm result has found. Various scientist has worked on Time-Series Model but the result they is up to limitation. Not only the work has been done on the traditional method, but also recent machine learning algorithm shows the better result such as ANN achieved better performance than those using only conventional indicators [14][15]. Even after achieving better result these systems have limitations in the area mainly based on the supervised learning.

In the learning task of stock prediction, it is more important represent target values by the successive relative changes in price since the previous time point than the absolute prices after a fixed time horizon which are generally adopted in conventional time series approaches based on supervised learning methods. In this case. Reinforcement learning can be the alternative algorithm method approach for stock market prediction.

A pure perennial neural network (RNN) classifier wasn't chosen for sentiment analysis as a result of it'd fail at distinguishing discriminating phrases occurring in numerous orders. The convolutional layer can fairly determine discriminative phrases in a text with a max-pooling layer [16]. Thus, the convolutional neural network (CNN) better captures the linguistics of text compared to RNN. Addressing the issues discussed in “Recurrent Convolutional Neural Networks for Text Classification” [16], We choose perennial convolutional neural network (RCNN) because the network. The RCNN accepts word embeddings that may be a result of text preprocessing because the input. The RCNN combination provides benefits of RNN and CNN. Max-pooling of the convolutional layer extracts the most effective illustration of the input. perennial nature of the network captures the discourse data to a bigger extent whereas learning word representations.

In this paper, scientist has proposed a novel hybrid approach based on the combination of the hodrick-prescott filter (HP) and Support Vector Regression algorithm (SVR). Its new hybrid for financial time series prediction using hybrid model. In this particular research scientist are predicting closing price because according to them its very important factor for investing the money into any stock. [17]

In this paper, scientists are approaching for long term stock market prediction using

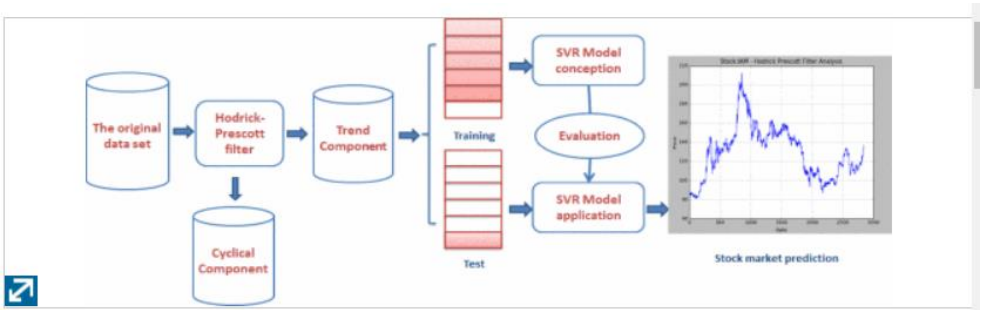


Figure 2.3

anomaly detection. Because in efficient market hypothesis the market always follows

random walk because bulk of trading volume follows some random distribution. Thus scientist thinks that there should be some random anomaly as if volume is not so random anymore. In such scenario long term predicting strategy is possible because stock price is no more random walk. Here they are studying using detected anomalies long term prediction is possible or not from analyzing historical data. [18]

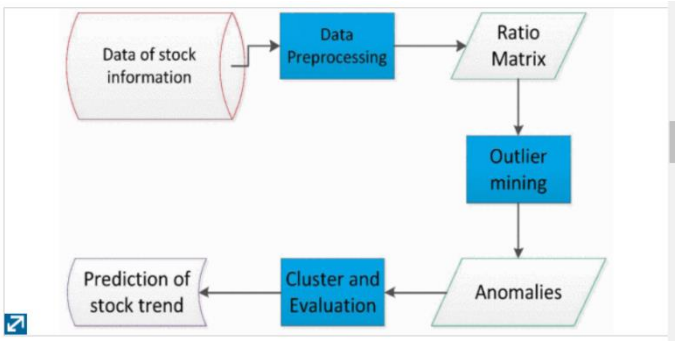


Figure 2.4

In this paper, scientists are saying that existing feed – forward ANFIS model is unable to cope to dynamic market character but can capable of static and recurrent type ANFI. The aim of this particular research is that the model that follows trajectory rather than attain fixed value. Scientists are predicting the closing price for next day based on analyzing the closing price of previous day. And proposed reliable RENFSM model can sustain to predict values with higher accuracy rates.

J.B Heaton has explained very in his research paper that there some problems while predicting the stock market price but it has some advantages of deep learning model over traditional machine learning algorithms method. And few of them help to avoid overfitting model easily and correlation data is also can be handled easily. [19]

In Torkil Aamdot thesis found that Support Vector Regression, Feed Neural Network and Convolutional Neural Network had comparable results and echo state network had noisy behavior. Not only this recently so many different types of algorithms and different types of neural network has been developed and still being developed. After all, having these developments finding suitable neural network for the data set is difficult. [20]

In this paper, according to Hengjian Jia noticed that LSTM works very effective on the stock market prediction and it has very decent RMSE value. Its values vary with different types of LSTM architecture and it can be improved. This study helps us understand the problem which arrives for time series data set and provided solution to tackle this problem. Scientist has successfully proved LSTM is the for time-series prediction. [21]

In this paper, scientists has proposed method named as Local Average Model (LAM) with turning points (TP).the average price and time are both important factor in the financial timeseries data because both are used for prediction of price in future. If the data is nonstationary nature then turning plays vital role to calculate the trend of financial time series. [22] [23]

In this paper, scientist has used ARIMA model to predict index series of Bangladesh. And he got very good result in that. He has concluded that ARIMA(3,2,1) is the best suitable

for their Bangladesh stock market index data set. And similarly for ARMA(3,1) [24]

This paper adopts the reinforcement learning algorithm to markov process problem of stock market prediction. For this stock market price process of reinforcement learning algorithm contains elements such state, action, reward, policy etc. and TD algorithm. [25], without completing the model environment reinforcement algorithm learns from raw experience like DP (Dynamic programming). The one main advantage of TD algorithm is that, Algorithm update estimates based in part on other learned estimates without waiting final output of program. [26]

Chapter 3: Research Aims and Objectives

Time series forecasting is done by analyzing the historical data and this data is used in the predictive model such as Time series algorithm, Deep learning, Artificial Intelligence and etc. which make the relationship between independent variables and dependent variables that help for prediction. Its not just about the applying the algorithm, programmer has to find out the most relevant attributes which contributes to strong hyper – parameter impact to make accurate model accuracy. Sock market price prediction is the demonstration of attempting to decide future investment of traders and organization's stock.

Research Question –

Comparison of Deep learning LSTM algorithm, Reinforcement algorithm with Time Series Algorithm approaches for stock market price prediction by analyzing google stock price.

Aim –

Helping in decision making for investor traders for stock market price using time series analysis

Objective -
To study and compare different machines algorithms such Deep learning and Time Series Algorithms. These algorithms have been trained on 15 years of google stock data to predict the future stock price.
The main objectives are:

1. Analyze the data set and identify the model
2. Model estimation and performance tuning for model
3. Calculate model accuracy.

3.1 Prediction Method

1. Deep Learning LSTM Algorithm
2. Deep Reinforcement Algorithm
3. ARIMA Model
4. ARMA Model

3.2 Proposed Approach

Data Mining is the process of finding out the trend pattern in the dataset whether is it belongs to the past transactional data or future trend. It gives idea/help to organization to take their decision effectively to make their business/project success. Basically 3 techniques are used in the data mining which are CRISP – DM, KDD and SEEMA.

But during past recent years so many changes has made, volume of data has increased in the field of Manufacturing Domain, Medical Domain Data, Engineering Applications, ERP system and Stream Analytics, etc. to handle these kinds of situations, we need some extended version of CRISP DM Model such as Extending CRISP-DP to Stream Analytics, CASP – DM (Context Aware Standard Process for Data Mining), Holistic Extension to CRISP-DM for Engineering Applications, etc. All these kinds of Methodologies will discuss in this paper. But to understand all these Extension, first we need to understand what CRISP-

DM is.

What is CRISP-DM? – Cross-Industry Standard Process for Data Mining is the most common methodological framework for designing, creating, building, testing, and deploying machine learning solutions. Built by more than 200 data mining experts, data mining tools and service providers. It is an industry tool which motivates the big corporates to realize the business objective in a better way. Fig 4 shows the six phases of the CRISP-DM.

CRISP-DM was lately created by four people those are Daimler-Benz (now Daimler Chrysler), Integral Solutions Ltd (ISL), NCR and OHRA of nascent data mining market. While data mining was being mounted there was no clear vision to data mining exists. At that time, a need of data mining technique was crucial which will standardize the industry and get help to run their data mining project through this they will get the best market to maturity and help to realize the good result from the data mining. (27)

In 2000, In Next – generation of CRISP-DM version 1.0 seen significant progress development in standardize the data processing model and expected to future expansions and improvement be made. Currently, industry experts are quickly accepting the six steps

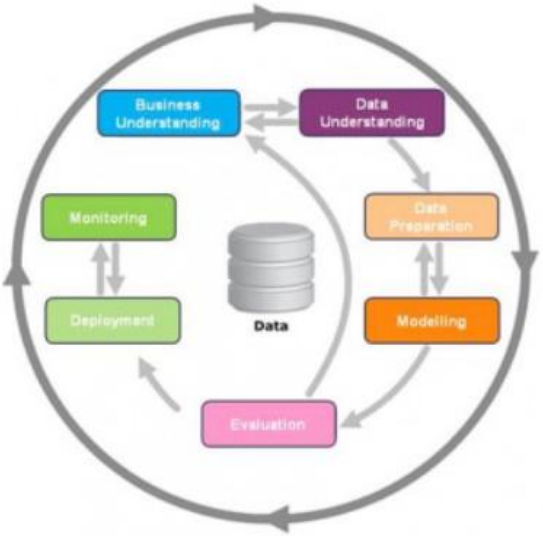


Figure 3.1

of CRISP-DM methodology. Which is Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.

Dissertation Proposed Steps are as follows:

Step 1: Business Data Understanding 1) Business objectives 2) Business importance 3) Sample data collection 4) Describe and explore the data

Step 2: Data Preparation 1) Transform the data to stabilize the attributes 2) Integrate and format data

Step 3: Data Modeling 1) Examine the data to identify potential models 2) Estimation and testing 3) Predict and forecast

Step 4: Data Evaluation 1) Evaluate results 2) Compare models

Chapter 4: Methodology

In this Research method we are discussing four types of methodology which are LSTM, ARIMA, ARMA and Reinforcement Learning. There are lots algorithm available in machine learning. Its has been noticed that easy machine learning algorithms gives astounding results as compare to the complex algorithm. It doesn't mean that easy algorithm is bad but it depends on which dataset which algorithm provides good result. We have followed this way of method to choose machine learning algorithms.

With the use of machine learning algorithms, forecast the time series unknown data is very technology. In the new recent era new deep learning, artificial Intelligence, Augmented Reality kind of algorithms being developed by researcher and its very challenging task for them. The experience on try-error on the algorithms helps in understanding in simpler way how machine learning can be used in forecasting. Behind all researches, researches have only aim that the predictive model they are developing it should estimate the value of unknown variable. In the time series model, three factors are very important which are time *(t)* as an independent variable, target dependent variable *(y).* and the final output of the model is the predicted value for *y* at time *t(y).*

Sample dataset collected from Kaggle in the Text file format. Then according to CRISP-DM method has applied to process the data, cleaning the data. Then Cleaned data it used train the algorithm. Its created by using python language. all discussed four algorithms in the first paragraph has applied on the google stock price data set. it has data since 2004 to 2017.

4.1 Sample Data Preparation

This paper Google stock price data has been used. Google is on e of the American multinational organization which has several products mainly it relates to the internet services. Such Google Chrome, YouTube etc. not only this but also google has cloud computing, Search Engine, online advertising Technologies, software and hardware. It is one of the top five companies in the world.

Dataset has 7 columns each column has different values, column named as Date, Open, High, Low, Close, Volume, OpenInt.

Date – Here date column is working as the index
Open – Open Column indicates the Open price of Stock with respected to the particular date
High - High Column indicates the High price of Stock with respected to the particular date
Low - Low Column indicates the Low price of Stock with respected to the particular date
Close - Close Column indicates the Close price of Stock with respected to the particular date
Volume – How many trades has been done on the particular day.

Stock market data set is available at the Kaggle.com or dataset can be downloaded from yahoo finance. Just recently new updates has got released by python -Jupyter where yahoo finance is the library through which data can be directly fetched. But in our study, data has downloaded from the Kaggle.com, to preprocess the data, it has read by the python and then it converted into the time series *(ts)* format.

Feature Selection

Feature selection is crucial thing while modeling in the machine learning algorithms. Because it which attributes contribute impact on the accuracy. Some attribute are not important means they don't make difference if they are considered in while modeling or not. Because it will not make any affect on the accuracy of model. If algorithms are used all the attributes in the modeling, algorithm will take time to learn the historical data as the not important attributes has also included hence taking correct attributes are very very important. It is done by the feature selection.

In machine learning algorithm, selecting feature is very challenging. But in case of time series data there is no feature selection part. It doesn't mean that feature is completely off limit, instead of that it needs to be used very carefully due to following reason.

- A time series model performs excellent than using feature selection.
- If predicted values will be used as the feature, it has very high chances to produce error to the target variable which turns to the biased forecast. Doing feature selection, model can learn some patterns which help in making the forecast model in better way.

4.2 Time Series

Time series is chronological sequence of data points, usually all data points are occurred in the same time intervals. Time series algorithm can be applied to any variable that changes over time.

Components of Time Series

Trend, seasonality, Noise is randomness, curve and the level are the components of time series. Before proceeding to explain what, these components are, its more interesting to know all time series data doesn't not contain these components such as audio files. They have periodic cycles instead seasonal components. But apart from that most of the business data has seasonal components such sales data in the second quarter.

Level: Level is referring to the mean values
Noise: Noise or Randomness will be in most of the time series data points, but they are not co-related with explained trends.
Seasonality: Seasonality components includes means regular and predictable variations in the data points and if they are strongly co-related with week, quarterly then seasonality components include in the time series
Trend: Trend means its showing the trend of data points its may positive or negative. Such as

Walmart sales trend whether it was upward i.e positive or downward i.e. negative.

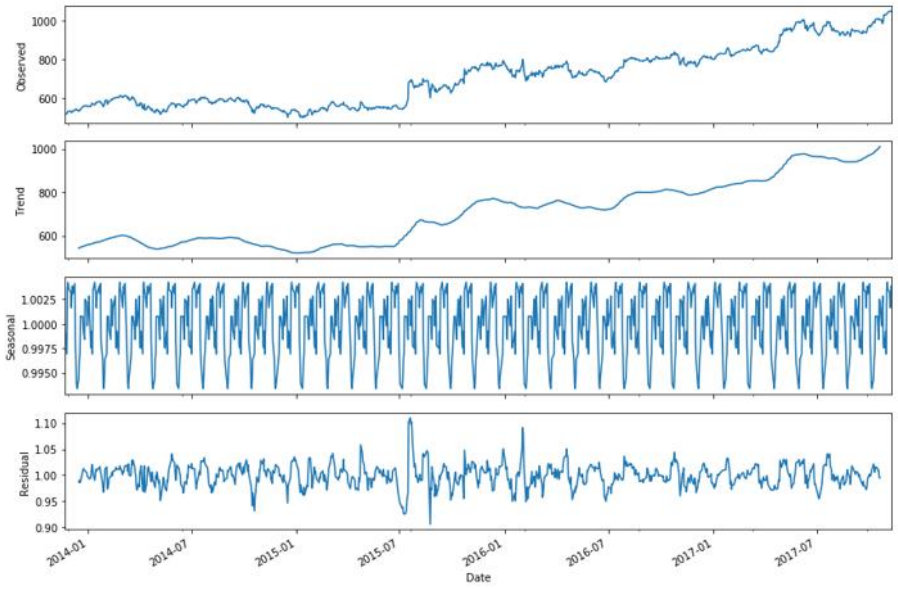


Figure 4.1

Cycle: Cycle refers to the repeating the cycles with some particular data point pattern.

4.2.1 AR Model

Autoregression model is the time series model which uses the input of previous time steps observations as input to predict the next time step value using regression. General equation of the AR model is based on the linear equation which is ***[yhat = b0 + b1*X1]*** where, yhat is the prediction, b0 and b1 are the coefficients got after model on training data. And X is the input value. The previous time step observations are taken as input as time series data is called lag variables.

For example, next time step(t+1) value can be predicted with given observations at the last two-time steps (t-1 and t-2). As a regression model. Which would look like as follows

$$X(t+1) = b0 + b1*X(t-1) + b2*X(t-2)$$

Linear regression model can be calculated manually using linear regression class in scikit-learn and manually mention the lag as input variables to use. Alternately stats model library has an autoregression model functionality with the using of this functionality model automatically selects an appropriate lag value using statistical tests. And train linear regression model. It is already available at the AR (Autoregression) class. Firstly this model is created after this to fit this model, python will call fit () to train the dataset. This is called AR Result Object. Once fit the model, prediction can be done calling function predict () for number of observations in the future. [28]

Similar type of Open loop feedback system is very really amazing thing in the autoregressive model which is type of control system. Main goal of this model is buried under some burden, mathematics but found some technique to build autoregressive model on the Wikipedia. There are two (1) & (2) equations with the help of these create expression (3). The result of this final expression having some intercept to be some part of autoregressive weights.

The goal of this function can be optimized with TensorFlow.

$$y_t = c + w_{t-1} * y_{t-1} \quad (1)$$

$$y_{t+1} = c + w_t * y_t \quad (2)$$

$$y_{t+1} = c \sum_{i=1}^{n} w_{t-i} + \sum_{i=1}^{n} w_{t-i} * y_{t-i} \quad (3)$$

4.2.2 Arima Model

Arima Model is statistical model which usually used for analyzing and predicting time series data. Arima is referred to as random walk which is fine tuned to reduce all residual autocorrelation. It is used for long-term trends and seasonality with exponential smoothing. Arima uses lags of the dependent variables and lags of the forecast errors as regressors

Construction of ARIMA Model

- Stationaries the series, if necessary, by differencing (& perhaps also logging, deflating, etc)
- Do lags of stationarized series or lags of the forecast errors need to be considered in the predicting equation or not it depends on study of the autocorrelations and partial autocorrelations.
- After fitting the model check residual diagnostics, particularly the residual ACF and PACF plots to study all coefficients are significant or not.
- Patterns that remain in the ACF and PACF my suggest the need for additional AR or MA terms.

ACF = Autocorrelation Function
PACF = Partial Autocorrelation Function

Terminologies in ARIMA
ARIMA model can be summarized into three factors which are P- The number of Autoregressive Terms, d = The number of nonseasonal differences, q = the number of moving average terms

P – The number of autoregressive terms

It allows incorporate the effect of past values into our model. Similarly learn the past values like if temperature for last 3 days were warm, it will assume that tomorrows temperature also be warm.

d – the number of nonseasonal differences

d is the number of nonseasonal differences needed for stationarity. This would be similar to say that it is likely to be same tomorrow's temperature if the difference of temperature was small in the last past three days.

q – the number of moving-average terms

q is the number of lagged forecast errors in the prediction equation of (MA part). This uses

the error values observed at pas time points of linear combination to set the errors.

These three integers (p,d,q) these parameter used in ARIMA models. Hence this called an "ARIMA (p,d,q)" model

ARIMA Filtering Box

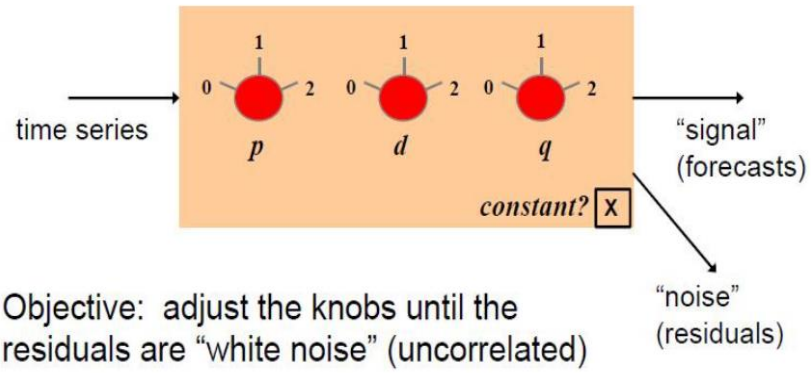


Figure 4.2

Working of ARIMA Model (p,d,q)
First apply differencing

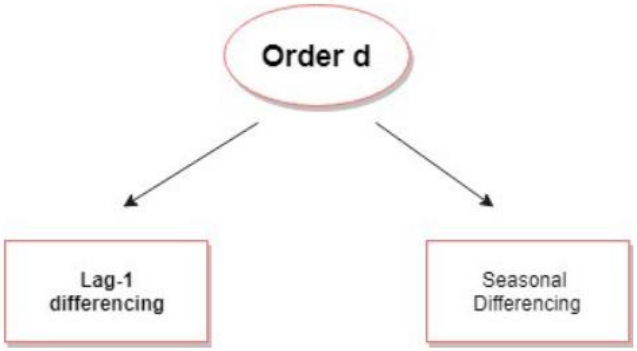


Figure 4.3

Then fit ARMA (p,q)

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

. . .

Lag-1 differencing:

$$y_t - y_{t-1}$$

Figure 4.4

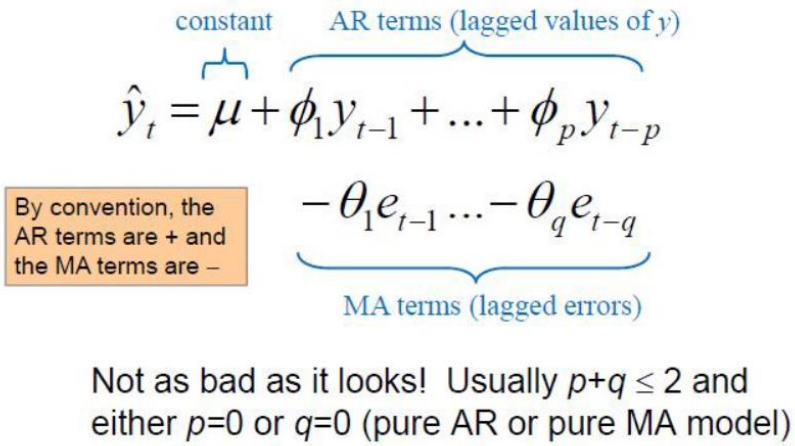
Used for removing trends. Order d = {1,1,0}
Order d = How many times to perform lag -1 differencing?
d = 1: one differencing
d = 1: perform differencing once (linear trend)
d = 0: No differencing

Arima Forecasting equation

- Let Y denote the original series
- Let *y* denote the differenced (Stationarized series)

First difference (d = 1) : yt = Yt
First difference (d = 1): yt = Yt – Yt-1
No difference (d = 2): yt = (Yt – Yt-1) – (Yt-1 – Yt-2)
= Yt = 2Yt-1 + Yt-2

Forecasting equation for y



Not as bad as it looks! Usually $p+q \leq 2$ and either $p=0$ or $q=0$ (pure AR or pure MA model)

Undifferencing the forecast

The differencing (if any) must be reversed to obtain a forecast for the original series:

$$\text{If } d = 0: \quad \hat{Y}_t = \hat{y}_t$$

$$\text{If } d = 1: \quad \hat{Y}_t = \hat{y}_t + Y_{t-1}$$

$$\text{If } d = 2: \quad \hat{Y}_t = \hat{y}_t + 2Y_{t-1} - Y_{t-2}$$

Advantages

Arima is the strong mathematical theory which enables it simpler to forecasting the Forecasting intervals. It is easy to noted many different parameters.

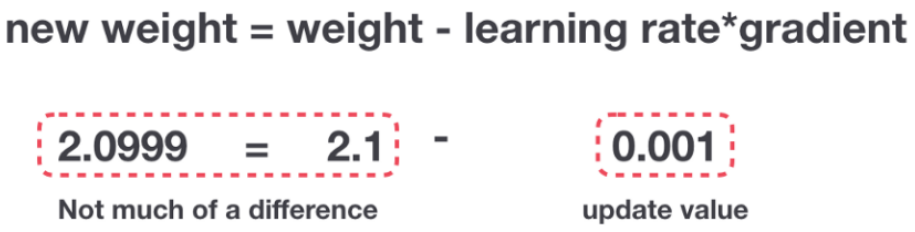
Disadvantage

Chances of getting model overfitted or misidentification because hard to recognize the coefficients and no explicit seasonal indices are present.

4.2.3 LSTM

LSTM model is been invented by the Sepp Hochreiter and JurgenSchmidhuber [29] which is the type of RNN (Recurrent Neural Network). Similar to the human brain RNN has short term memory which is called as LSTM so that output is related to the present and past. as this memory is short so it doesn't store the things for long term as like human brains. As LSTM does not have the property to store information for longer time to process it, the basic unit of LSTM network will overcome the limitations of LSTM of gradient vanishing problem and its been widely used in many organizations.

Gradients are usually update neural network weights but during back propagation, recurrent neural network has problem of vanishing gradient problem. The main important things that need to be understand is that when gradients shrink, the vanishing gradient problem happens as it back propagates through time. but it doesn't make any differences in learning if gradient value is very very small.



Gradient Update Rule

One thing happened while building algorithm is that, on which layer get small gradient update, it stops learning generally all these layers are belongs to the earlier layers. Recurrent neural network forgets the things which are seen longer in longer sequences. That's the propose of the Long short-term memory and GRU being invented to over come the problem of the short-term memory loss.

There are five gates has used in the LSTM model which help to learn the which data sequence is important or not and that are those need to be kept or not. With the help of the that gate passes the information which are essential to long chain sequences to make predictions. Based on the recurrent neural networks almost all state art results getting with those two networks. LSTM model can be used in the speech reorganization, speech synthesis, and text generation. These

algorithms are also used to captions the videos

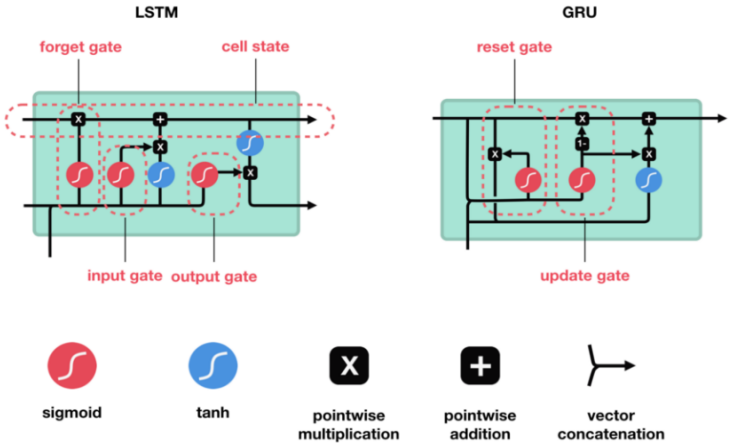


Figure 4.5

Review of Recurrent Neural Networks

To get better idea how LSTM achieves the learning, let review the RNN neural network. RNN get first words and transformed in to the machine – readable vectors. As like shown in below diagram then it posses sequences of vectors one by one

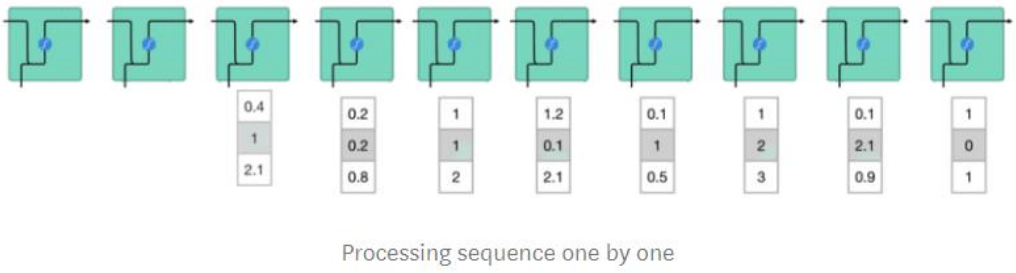


Figure 4.6

How it processes the information, while processing previous hidden state passes the relevant

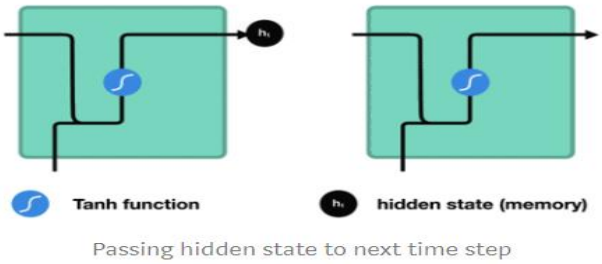


Figure 4.7

information into next sequence step. Hence generally hidden state act as the neural networks memory. It also holds the information of existing previous hidden state network before seen it. In every algorithm model, hyperparameter tuning is very crucial thing to increase accuracy. Without hyper tuning algorithm works as biased. Hence, we need to understand that how to calculate the hidden state. Firstly, vector is needed to be formed by combining input and previous hidden state. Now the information which is available in the vector on the current input and previous inputs. After that to active the function, vector goes through the tanh activation, and the output is the new hidden state, or the memory of the network.

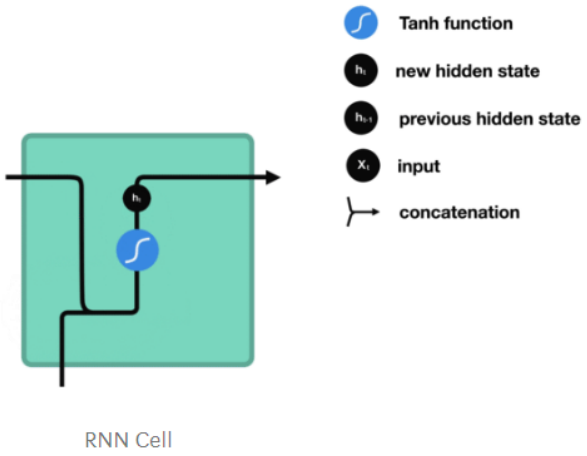


Figure 4.8

Tanh activation

Tanh activation is used to help regulate the values flowing through the network. The tanh function squishes values to always be between -1 and 1.

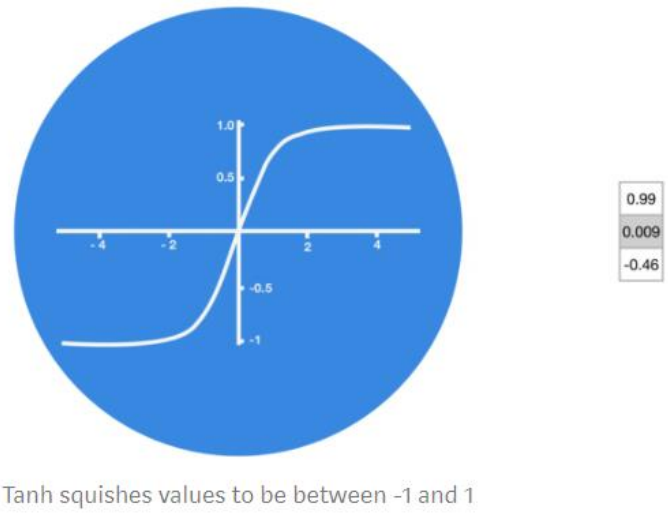


Figure 4.9

Due to various math operations, when vectors are goes through the neural network, it get so many transformations. So it can be calculated that value that continuously multiplied by lets say 3. Then some values can explode and become astronomical, because other values are getting insignificant. Thanh function assures that values will be between -1 and 1, hence it very easy to regulate the output of the neural network. In below image it is clearly mentioned that same values above remain between the boundaries allowed by the tanh function.

Core - Concept

The core concept of the LSTM's are the cell state, which has various gates and it behaves like mode of transport check post to check motor if they are ok carry limited amount of carriage then they are allowed. Similarly, in case of LSTM gate information is checked before proceeding to next sequence chain. Most of the scientist has mentioned in their research that the cell state carries the relevant information throughout the complete processing of sequence. So, the information from its previous time step they make its own way to processing for later time steps., reducing the effects of short – term memory. As the cell starts their processing the information, relevant information may get adds or removes from the cell state via gates. The gates are the different neural network that are responsible to take decisions on the information which has to be allowed or not. The gate has full authority to learn what information is relevant to keep or forget during training.

Sigmoid

Gates have sigmoid activation function. Sigmoid is just almost similar like to the tanh activation function. Incase of tanh activation values are stays between -1 and 1, but in sigmoid activation values are compressed and keep stays between 0 and 1. Which is most fabulous thing to update or forget data because any number multiplied by 0 is zero. Which causes values to be disappears or be forgotten. Any number multiplied by 1 is the same value hence the sam values stays same or is kept same. Due to such terminology the network can understand what is important and which need to be kept and which are not. That means forgotten.

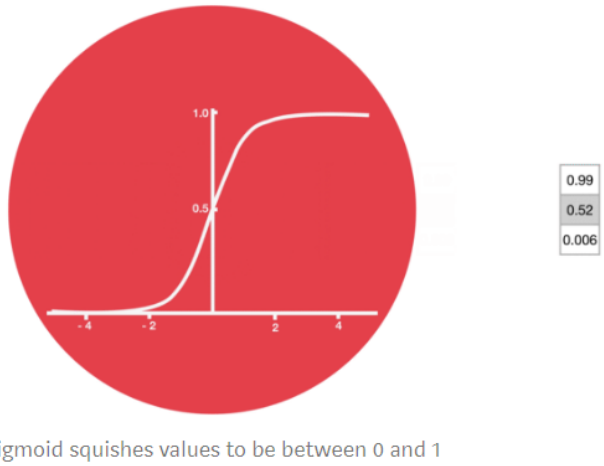
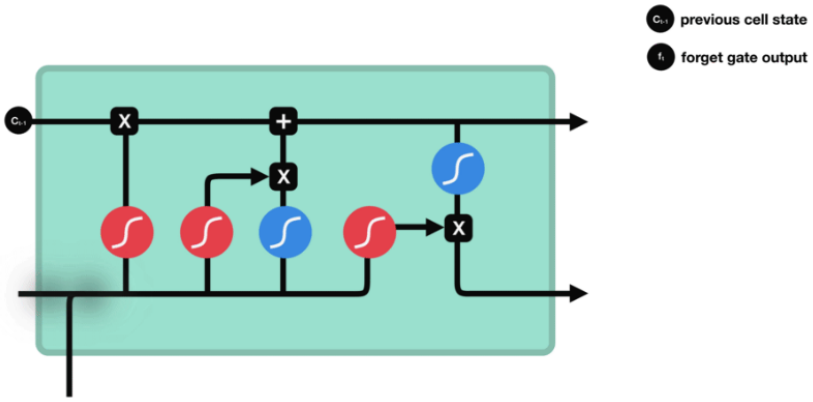


Figure 4.9.1

There are three gates forget gate, input gate and output gate with the help of these gates LSTM regulate the information flow. Below diagram and concept are explaining the deeper concept of this gates.

Forget Gate

Forget gate is one of the gates used in the LSTM network. It helps model to take decisions on which information is relevant so it need to be kept or need to be thrown. While modeling there lots of hidden layer is made to increase the accuracy and according to the dataset through which many information passed on which forget gate takes crucial decision. Every layer passes this information from previous layer to current input through the sigmoid activation function. Then values come out between 0 and 1. If the value is closer to 0 means forget and the information value is closer to 1 means to keep

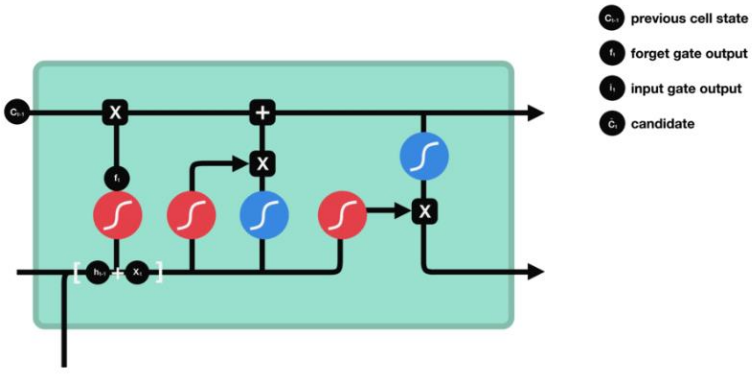


Forget gate operations

Figure 4.9.2

Input Gate

If cell state has to be updated, then input gate is responsible for that. Firstly hidden state is passed and current input into the sigmoid activation function. Then sigmoid function take decision which values will be updated by transforming output of the values gets between 0 and 1. If output is 0 that means it not important so it can be thrown away and if output is 1 that means it is important so it can be kept for next hidden layer as input. Not only this but also through tanh function information is passed to compress the value between 0 and -1 to help regulate the network. After all this operation tanh function and output from sigmoid function is multiplied and output of these multiplication will decide which information crucial to keep from tanh output.



Input gate operations

Figure 4.9.3

Cell state

Now model has sufficient information to calculate the cell state. Firstly cell state is multiplied by the pointwise forget vector. If vector is multiplied by values near to zero then vector has very chances of dropping values in the cell state. After this operation output of input gate do the pointwise addition which update the cell state to new values that neural network finds relevant. Which ultimately gives new cell state.

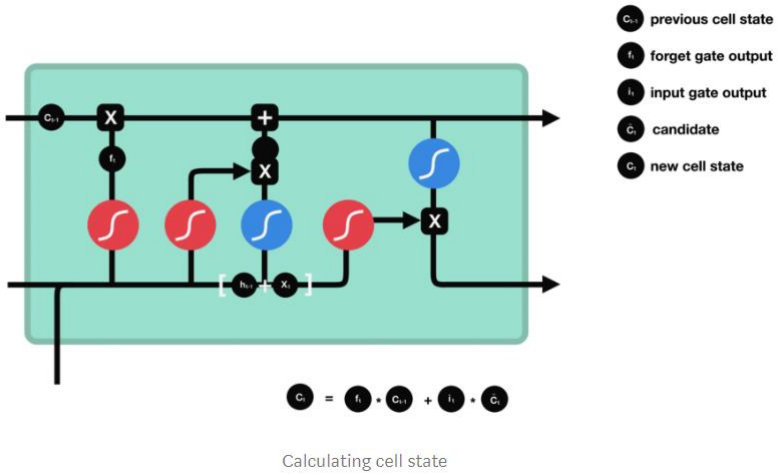


Figure 4.9.4

Output Gate

Lastly output gate will be explained here. It is responsible to take decisions what is the next hidden state will be. The hidden state has information of previous which is used also used for predictions. Like all the input gate first previous hidden state and current input into sigmoid function. Then pass the newly occurred cell state to the tanh function. Multiply the output with sigmoid output to decide what information need to carry by the hidden state and output is hidden state. The new cell state and new hidden is pushed to next time step.

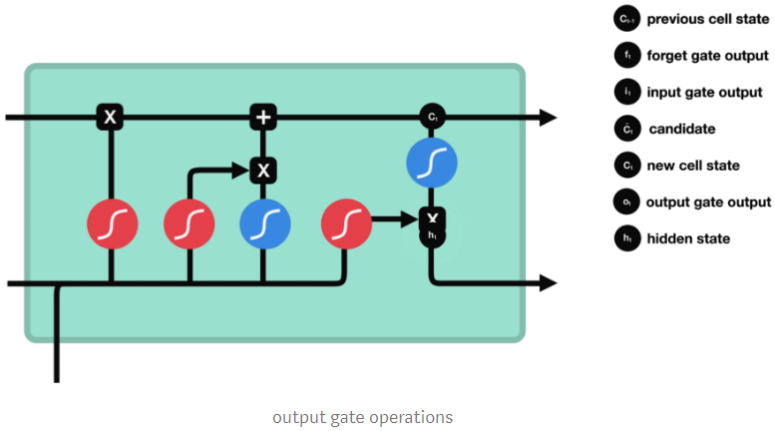


Figure 4.9.5

Forget gate take decisions what information have to keep from previous state. The input gate decides

what kind of information have to add from the current step. And the output gate finds what will be the next hidden state.

4.3 Software and Packages

In this research paper all programming has been done on the python why python is very important? Because it is easy to learn, scalability, choice of data science library, python community, graphics and visualizations. Each factor will be discussed below.

Easy to learn

Python is very easy download, its open framework anyone can download this from internet. And lots interface is available to start coding on that. It is very easy to learn as well lots programming books are available to learn. On another site Udemy, YouTube is one of the best way to learn. When python compares with other data science language like R, python is promoting to takes less time learn by millions of users by promoting easy to understand syntax and which is most difficult phenomenon in any programming.

Scalability

When it comes to python to compare with other languages like R, python has already created lead among all other languages. And it is much faster than other languages like MATLAB and STATA. YouTube is one of most inspiring web applications which was made on the other language but after it has been migrated to python by influenced by the scalability. Python has also lies in the flexibility that it gives to solve problems. Python has come good for different usages in different industries and for rapid development of applications of all kinds.

Choice of data science libraries

There are lots of libraries available in python. There variety of data science/data analytics libraries are made to help out the people which are working in data science field on python. Pandas, StasModels, NumPy, SciPy, and scikit – learn are some of libraries from the python are very famus in the field of data science community. Python library developer has not stop with those libraries only, they keep making libraries which help people to deals with new difficulties and new approach.

Python Community

One of the reasons for excellent rise of python is because of huge number of python community. As and when python reach to the data science community, tried to maximum number of volunteers are creating data science. There lots of python community's people working on different web sites, linkdin group and etc. where all user gets their solution if they have any problems in coding. stack overflow is one of the famous communities where all most all time people get help for their queries

Graphics and Visualization

Python comes with variety of visualization libraries option and usually they are compatible with other libraries also. Matplotlib provides the solid foundation with other libraries like seaborn, pandas plotting, and ggplot have been built. Visualization packages give some attracting data exploration where user get proper idea of data understanding. Python has such visualization libraries

through which programmer can make interactive dashboard and be deployed at web as well.

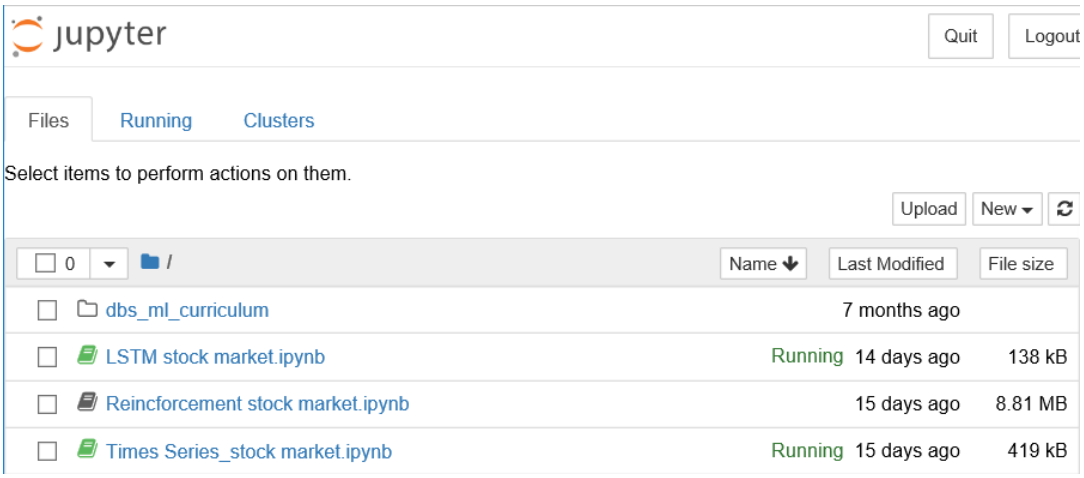


Figure 4.9.6

Packages and Libraries:

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

import matplotlib.pylab as plt
%matplotlib inline
from matplotlib.pylab import rcParams
rcParams[‘figure.figsize’] = 15, 6

from sklearn.preprocessing import MinMaxScaler
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM, GRU
from keras.layers import Dropout

from statsmodels.tsa.arima_model import ARIMA
from math import sqrt

import math
from sklearn.metrics import mean_squared_error, mean_absolute_error
```

Number of Files: 1 CSV file which contain google stock price data

Number of records: CSV files contains the stock price data since 2004 to 2017

Chapter 5: Data Analysis

5.1 Exploratory Data Analysis

Exploratory Data Analysis is critical process of initial data analysis, investigations on data to find out the patterns, to identify anomalies, to do hypothesis with the help of summary of statistics and graphical representations.

It is very good techniques to learn and understand the data first, what kind of factor has been included it, and try to find out the as many as insight from it. Usually it is said by the scientist that EDA is al about making the data in hand, before getting them dirty with it. With the help of EDA, how attribute are co-related to each other, it can be calculated. Let's take an example of heat map. After creating heat map, there are so many small squares are generated making together it become big matrix. But important thing is to notice in that dark shades represents the positive co-relation while lighter shades represents negative co-relation.

In this research paper, auto co-relation technique has been used, which clearly show that all data lies on the single line, so it can be concluded that this is positive co-relation.

Step 1: Generate questions about data
Step 2: Search answers by visualizing, transforming, and modeling the data
Step 3: Use what is to learn to refine questions and if required then generate new questions

```
from pandas.plotting import lag_plot
plt.figure(figsize=(12,8))
lag_plot(data['Open'], lag=1)
plt.title('Google stock Autocorrelation plot')
plt.grid(True)
plt.legend();
```

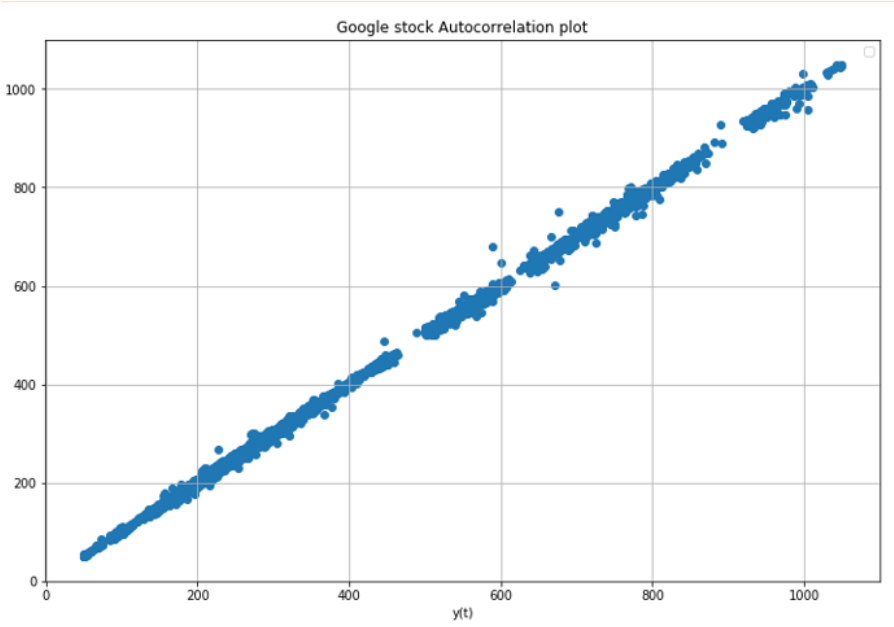


Figure 5.1

The above auto co-relation plot is showing relation between the target values with the input values. where in x-axis showing the output variable in this case output variable is Open price of stock market and y-axis is the date which is input variable.

5.1.1 Sample Data Analysis

After sample data analysis, another task is modeling the sample data analysis and visualization to check weather is there any null values is there, or any zero values are present in the data? Because the if these two values persist in the data set, after generating the predictive model it will not provide accurate prediction. There are few techniques has been invented to deal those values but, the stock market price dataset which is downloaded from Kaggle it does not need any kind of techniques to apply because this data is proper. There is no zero and null values. it is uniformly distributed so it will give biased prediction after generating predictive model. how data points have spreaded over year from 2004 to 2017 is mentioned in the below graph.

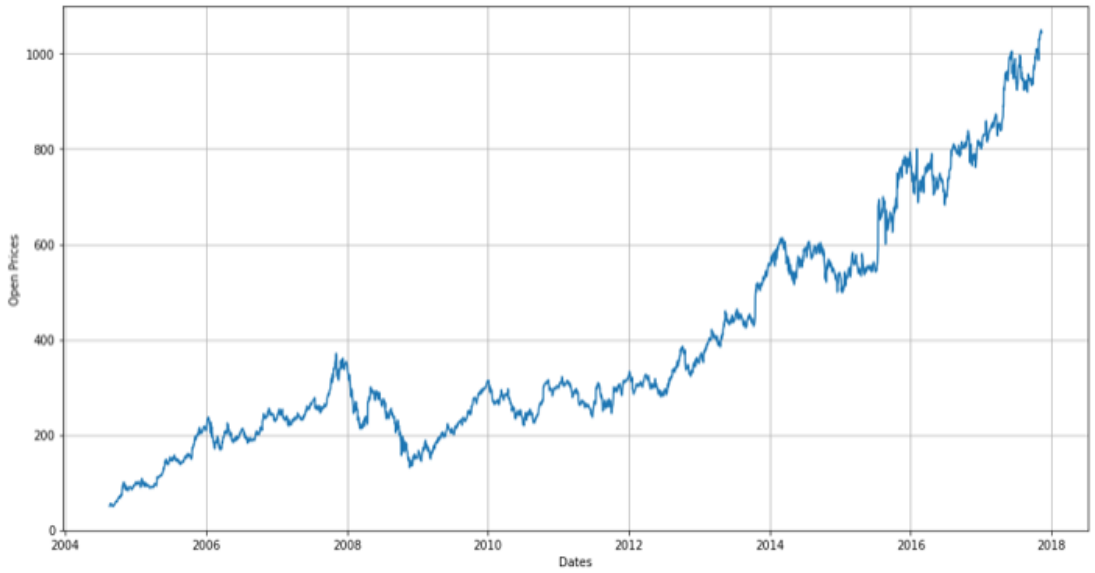


Figure 5.2

From above figure it is clearly shown that data has plotted for year since 2004 to 2017. On x-axis dates are mapped and, on the y-axis, open price of google stock are mapped. The scale of y -axis is starts from 0 to goes up to 1100. This line graph is clearly shown the upward trend throughout all years. But there are some up-down trend as well between year 2008 to 2010. First four year of google stock was very good. Stock price for those years were tremendously increasing. But after 2008 there was deep drop in the stock price till 2009. Before 2008 stock price rate touched to almost 400 after until 2008 prices were dropping. 200 stock price has been noted in year 2009 and it seems to be considered as lowest price of google. But after 2009 it started again performing well. The average price was 300 between 2010 to 2014. There were not so much trend has observed in those years. Prices were almost steady for those years. Little bit up and down trends were there but it was like not so noticeable, the attracting trend comes in picture when prices starts increasing after 2014. There was rapid growth has been observed when price touched to 800. the way of moving stock prices was like exponential growth its clearly seen in the graph. And in 2017 stock price crossed 1000. Google has done more 900% growth from year 2004 to 2017.

5.1.2 Error Value Analysis

After building model the first and important thing has to see that how accurate is model. to get understand this, cost of error should be discussed and understand. What is cost error? How it impact on the model? what is the role of cost error? How important is that?. Its very stupid thing to assuming something which is true but in actual it is nor, or assuming it is false when it is true or mot guessing at all. For an example, it might get sometime cheaper lose the customer than retaining it. Means sometimes investment is more to retain the customer and when profit is calculated at the end is less than losing the customer. or cost of vaccination is need for such people who are sick than the increasing number of getting sick people. This kind of cost/benefit analysis methodology gives idea to identify the probability of getting wrong and correct future predictions. Thus its very important to calculate the error measurement.

To get better idea how model is working, weather it is trustable, or not should it need to be considered in the future for real time environment data or not. Hence it is very important know how much model is perfect to cover all these kind of hypothesis by calculating the error values. there are variety of methods calculate the error values. confusion matrix is used to display multiple type of error measurements so data scientist can determine if the model is performing well or not. The most popular error measurements are as follows:

- ROC Area
- Accuracy (ACC
- Mean Absolute Error (MAE)
- Specificity or True Negative Rate (TNR)
- Root Mean Squared Error (RMSE)

ROC Area (ROC AUC)

In world war two, ROC was originally used for radar object detection. The ROC or "Receiver Operating Characteristic". Generally its ranges from 0 to 1, and if it is higher then it is better. The ROC area is measure of the area which comes under the curve produced by graphing the ratio between TPR and FPR with values ranging from 1 - .8 is amazing to good, .8 - .6 is fair to poor and if it comes below that is not referred as random chance.

Accuracy (ACC)

Usually accuracy range is from 0 to 1, and it referred better if value is higher. It simply calculated by taking ratio of correctly predicted observed values to the total number of observations.

$$ACC = (TP + TN)/(TP + FP + FN + TN)$$

Instinctively one is considered to be great measurement but it in actually is shows very less information about false positive and negatives.

Mean Absolute Error (MAE)

MAE is one the method of error calculating. Which ranges from 0 to infinity and it is considered to be lower is better. Is almost similar like to RMSE. It is calculated by the taking the average of absolute differences of residuals instead of taking squaring the difference of residuals and taking

the square root of the result. MAE produces the positive numbers only and is less reactive to large errors but can show nuance a better.

Specificity or True Negative Rate (TNR)

True negative or specificity is used to measure the proportion of negatives that are correctly observed in the model. e.g the percentage of healthy people are correctly predicted by the model. it is very useful when it comes to measure the accuracy of the model.

TNR = TN / (TN + FP)

Root Mean Squared Error

RMSE stands for stands for Root Mean Squared Error. Which has range from 0 to infinity, if the range is as much as lower is considered as lower. RMSE is also called Root Mean Square Deviation (RMSD), it used to measure the absolute average magnitude of error by quadratic – based rule. Technically is calculated by taking residuals of the values then squaring it, averaging all the results and then taking square root of the average. But what is residuals it need to be understood. Residual means difference between the predictive model and actual data. Because this phenomenon the output of this product is always positive.

Because squared is taken before taking the averages of values, the affect this concept is greatly amplified and should be used of those kinds of errors are important to identify.

RMSD can be normalized by taking the mean or range in order to compared between models of different scale. Usually it is expressed as percentage and noted as NRMSD or NRMSE

Here, in this research paper, RMSE value is comparing with all models. Confusion matrix can not be calculated because data is continuous so on the basis of the RMSE values, which model is working well is being concluded. RMSE values of the different algorithm model is as follows.

RMSE value of Auto Regressive Model

```
: from sklearn.metrics import mean_squared_error, mean_absolute_error
import math
mse = mean_squared_error(test_data['Open'], predictions)
print('MSE: '+str(mse))
mae = mean_absolute_error(test_data['Open'], predictions)
print('MAE: '+str(mae))
rmse = math.sqrt(mean_squared_error(test_data['Open'], predictions))
print('RMSE: '+str(rmse))

MSE: 78.49759192800558
MAE: 6.4254482287926065
RMSE: 8.85988667692796
```

Figure 5.3

Here incase of stock market prediction autoregressive model is giving RMSE value is 8.85 and it considered as very result, it almost more than 85% accurate model than LSTM. In this data set data from 2015 to 2016 is train dataset and from 2016 to until 2017 is test dataset. Green line indicates the train data and red line shows the test data and blue line which is overlapping on the green line is the predicted value produced by the predictive model. and it is clearly seen that predictions are

quiet good when it compares with the test data.

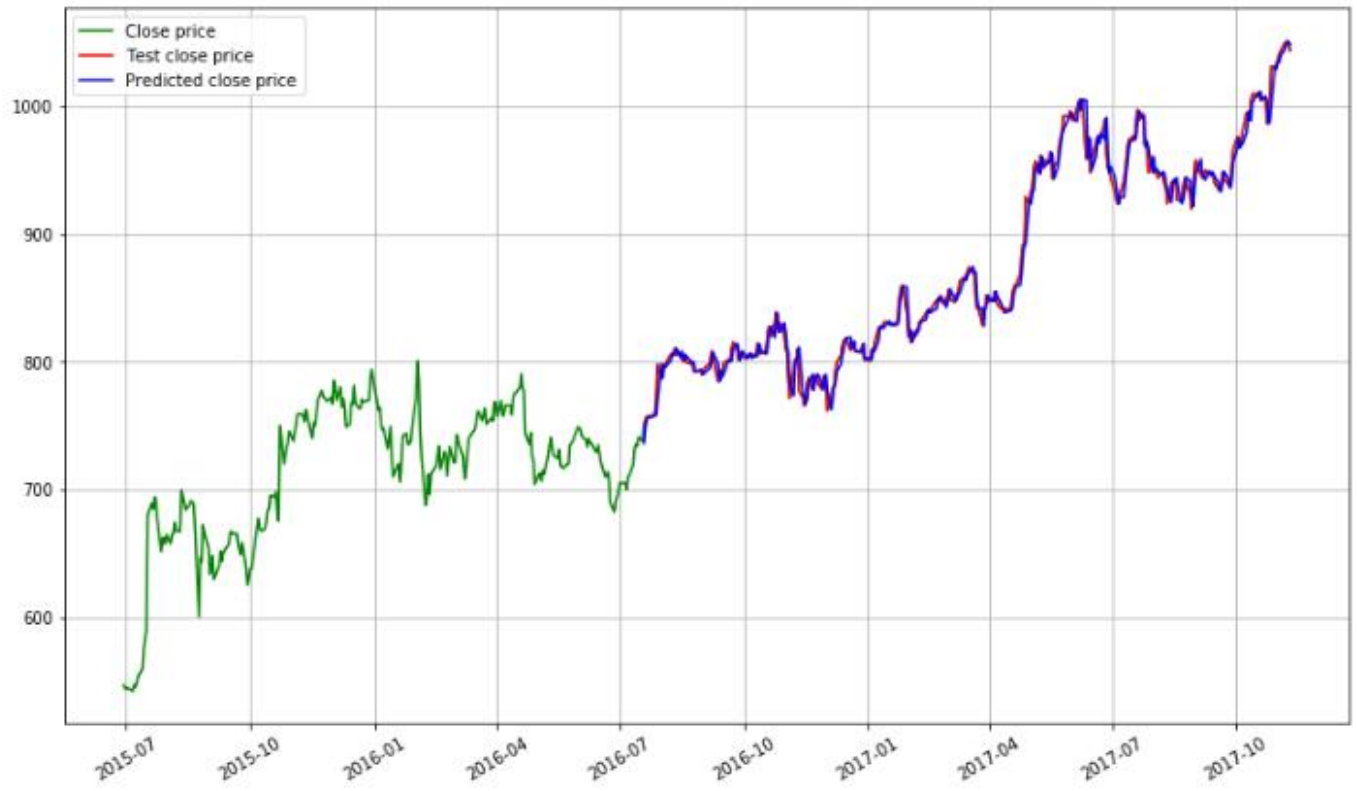


Figure 5.4

RMSE value of ARIMA Model

RMSE value of the ARIMA model and AR model is almost similar. They are working almost similar on google stock market data set. when RMSE value both model have compared it has observed both model have same accuracy value. But this is not case for LSTM, it has higher RMSE value and working poorly than both time-series model. which has described in another paragraph.

```
# report performance
mse = mean_squared_error(y, predictions)
print('MSE: '+str(mse))
mae = mean_absolute_error(y, predictions)
print('MAE: '+str(mae))
rmse = math.sqrt(mean_squared_error(y, predictions))
print('RMSE: '+str(rmse))
```

```
MSE: 78.41952836731198
MAE: 6.249253612247241
RMSE: 8.855480131947221
```

Figure 5.5

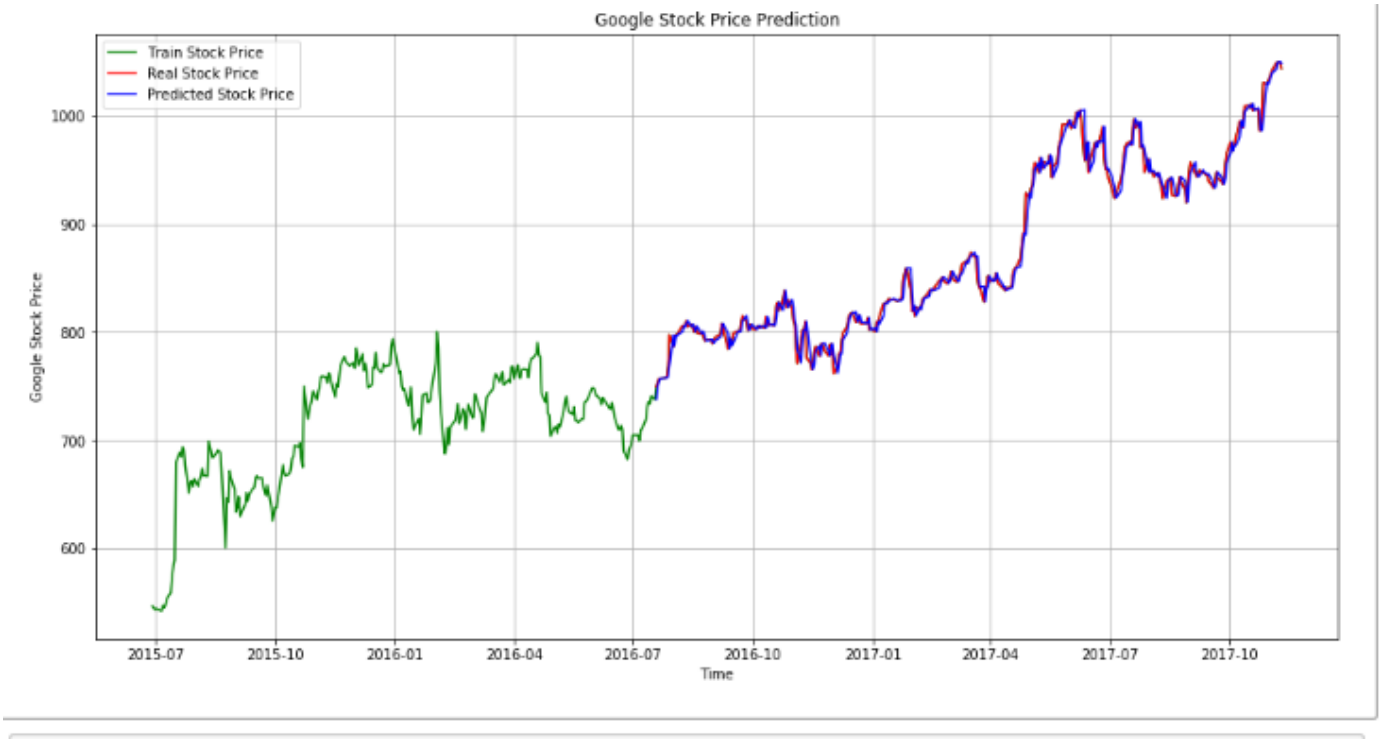


Figure 5.6

RMSE value of LSTM model

```
from sklearn.metrics import mean_squared_error, mean_absolute_error
import math
mse = mean_squared_error(real_stock_price, predicted_stock_price)
print('MSE: '+str(mse))
mae = mean_absolute_error(real_stock_price, predicted_stock_price)
print('MAE: '+str(mae))
rmse = math.sqrt(mean_squared_error(real_stock_price, predicted_stock_pric
print('RMSE: '+str(rmse))
```

```
MSE: 1076.2833976521217
MAE: 26.38239567008561
RMSE: 32.806758414267655
```

Figure 5.7

The value obtained for LSTM model is 32.80, even it is not considered as too bad accuracy. It can say that model has predicting almost more than 60% values are accurate. Red line indicates the test data set and which are true values of stock price. And blue line indicates the predicted values by the

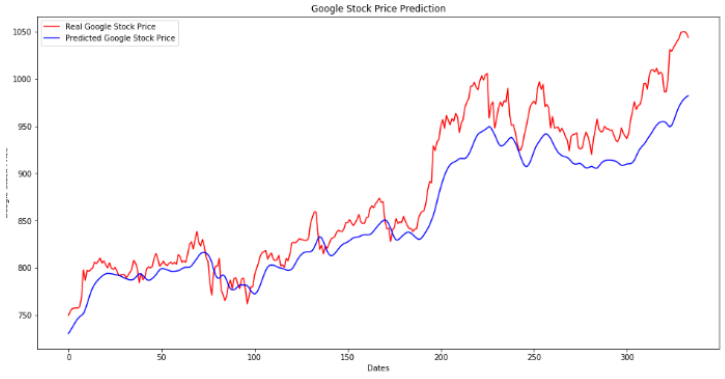


Figure 5.8

predictive model which is built by using LSTM algorithm. It clearly seen from the graph that predicted values are properly following the up-ward and downward trend occurring in true test values. and this model is also can be considered in the future for further updating. If compares three models based on RMSE values ARIMA and AR model has high accuracy rate than LSTM model. both time-series models are 20% more accurate than LSTM predictive model.

Chapter 6: Conclusion

This study has illustrated the excellence procedure of ARIMA model, AR Model, and LSTM model for stock market price forecasting. The detail examination of these 3 Models are unrevealed that stock data has trained by all three different algorithms. And the results that got from the Models through which it can be concluded that forecasting of stock market price for future reference are quite important when see the results of those models. The results are quite satisfactory. This can be helpful for trader to take decision about their investment money. With the outcomes achieving from Time – Series Algorithm are very well accurate than LSTM model.

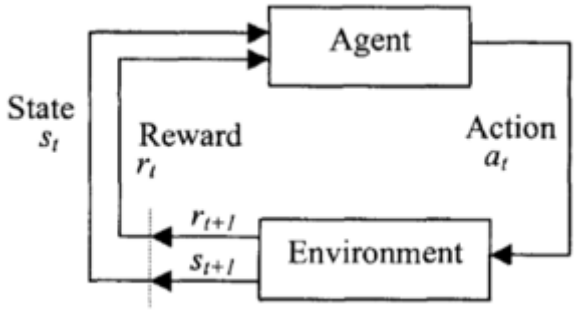
The Result got from the ARIMA Model is impressive, with RMSE value 8.8594, whereas other model that has built which is AR model with getting RMSE Value 8.8554. Both RMSE value has slightly difference. But they are almost similar because if mapping of forecasting value of test true actual values are compared, both has same up and down trend with similar value. On other hand LSTM is working with RMSE value 32.8067 which is also not considered as very bad RMSE value. From the research and study, it can be said that with the help of Deep Reinforcement Learning Algorithm Prediction could get better. How Reinforcement learning can work better, what are the impressive feature it has which is clearly distinguish this algorithm by any other time series algorithm has discussed in the Future work.

If there is strong seasonal pattern in the data, ARIMA and AR model are considered as trustworthy model, whereas the LSTM model is also capable to find out the type seasonality in the data and decompose the time series based on the type of seasonality and accurately predicted for all the data set. when all these three statistical models are compared, seasonal data is better handled by ARIMA and AR than LSTM Model which suggest that, ARIMA will able to forecast the data with strong pattern.

While comparing RMSE value for each model, it has been concluded that statistical performance of ARIMA and AR is differed from RNN LSTM method. Because Time-Series algorithm method is more suitable than statistical models in forecasting the stock market price.

6.1 Future Work of Reinforcement Learning

Reinforcement learning is a computational approach to understanding and automating goal-directed learning and decision making. It has very different approaches by its importance on the individual from direct interaction with its environment as shown in figure.



At every distinct time step t, the agent senses this state's, chooses a current action a,, and performs it. The atmosphere responds by giving the agent an award rt+1 = r (st, at) and by manufacturing the succeeding state st+1 = Delta(St, at). Here the functions S and r area unit a part of the atmosphere and don't seem to be essentially renowned to the agent. In MDP (Markov call process), the functions Delta (St, at) and r (st, at) rely solely on this state and action not on earlier states or actions. The task of the agent is to learn a policy, Pie : S ---> A, wherever S is that the set of states and A is that the set of actions, for choosing its next action supported this determined state s; that's, pie(st)=at associate degree best policy could be a policy which will maximize the attainable reward from a state, known as price, v(Pie) (s), for all states. [30] could be a typical definition of this:

$$V^{\pi}(s_t) = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \ldots$$
$$= \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \cdot \qquad (1)$$

Here y (0 < y < 1) is a constant that determines the relative value of delayed versus immediate rewards. V(m(s)) the possible cumulative reward achieved by following an arbitrary policy n from an arbitrary initial state. Then an optimal policy, n', is defined as follows:

$$\pi^{*} = \arg\max_{\pi} V^{\pi}(s), (\forall s) \cdot \qquad (2)$$

This work proves the thought that reinforcement learning is accustomed trade stocks. Stock trade isn't presently best solved with reinforcement learning, however the thought of a computer having the ability to get revenue simply by commercialism stocks is encouraging. Such non-deterministic issues will solely be solved with neural networks. The system built for this project works with the stocks of 1 company. This design may be scaled to require advantage of the stocks of multiple firms. Scaling this project would need coordination among multiple networks. A master network may be trained to leverage the predictions from individual company networks. The master would think about the actions foreseen by the networks and select among them the best actions it will perform with the resources it's.

References

1. Khare, K. *et al.* (2017) 'Short term stock price prediction using deep learning', in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*. *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pp. 482–486. doi: 10.1109/RTEICT.2017.8256643
2. Cheng, L., Huang, Y. and Wu, M. (2018) 'Applied attention-based LSTM neural networks in stock prediction', in *2018 IEEE International Conference on Big Data (Big Data)*. *2018 IEEE International Conference on Big Data (Big Data)*, pp. 4716–4718. doi: 10.1109/BigData.2018.8622541.
3. R. Gupta., "Emerging Markets Diversification: Are Correlations Changing over Time" International Academy of Business and Public Administration Disciplines (IABPAD) Conference, Orlando, 3-6 January 2006
4. B. G. Malkei, "A Random Walk Down Wall Street" 7th Edition, W. W. Norton & Company, New York, 1999
5. Yang, Z. *et al.* (2016) 'Stacked Attention Networks for Image Question Answering', in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21–29. doi: 10.1109/CVPR.2016.10.
6. Selvin, S. *et al.* (2017) 'Stock price prediction using LSTM, RNN and CNN-sliding window model', in, pp. 1643–1647. doi: 10.1109/ICACCI.2017.8126078.
7. Tiwari, S., Bharadwaj, A. and Gupta, S. (2017) 'Stock price prediction using data analytics', in *2017 International Conference on Advances in Computing, Communication and Control (ICAC3)*. *2017 International Conference on Advances in Computing, Communication and Control (ICAC3)*, pp. 1–5. doi: 10.1109/ICAC3.2017.8318783.
8. 'J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecast-ing,"'.
9. *Bulk Price Forecasting Using Spark over NSE Data Set* (no date) *ResearchGate*. Available at: https://www.researchgate.net/publication/303953824_Bulk_Price_Forecasting_Using_Spark_over_NSE_Data_Set (Accessed: 15 August 2019).
10. Box, G. E. P. (1994) *Time Series Analysis: Forecasting and Control*. Prentice Hall.
11. Batres-Estrada, B. (2015) 'Deep learning for multivariate financial time series', in.
12. Ding, X. *et al.* (2015) 'Deep Learning for Event-Driven Stock Prediction', in *IJCAI*.
13. Baumann, M. H. (2018) 'Beating the Market', in.
14. Kuo, R. J., Lee, L. C. and Lee, C. F. (1996) 'Integration of artificial neural networks and fuzzy Delphi for stock market forecasting', in *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929)*. *1996 IEEE*

International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929), pp. 1073–1078 vol.2. doi: 10.1109/ICSMC.1996.571232.
15. Baba, N. and Kozaki, M. (1992) 'An intelligent forecasting system of stock price using neural networks', in *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*. *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, pp. 371–377 vol.1. doi: 10.1109/IJCNN.1992.287183.
16. Lai, S. *et al.* (2015) 'Recurrent Convolutional Neural Networks for Text Classification', in *AAAI*.
17. Ouahilal, M. *et al.* (2016) 'Optimizing stock market price prediction using a hybrid approach based on HP filter and support vector regression', in *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)*. *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)*, pp. 290–294. doi: 10.1109/CIST.2016.7805059.
18. Zhao, L. and Wang, L. (2015) 'Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm', in *2015 IEEE Fifth International Conference on Big Data and Cloud Computing*. *2015 IEEE Fifth International Conference on Big Data and Cloud Computing*, pp. 93–98. doi: 10.1109/BDCloud.2015.19.
19. Heaton, J. B., Polson, N. G. and Witte, J. H. (2017) 'Deep learning for finance: deep portfolios', *Applied Stochastic Models in Business and Industry*, 33(1), pp. 3–12. doi: 10.1002/asmb.2209.
20. Aamodt, T. (2015) 'Predicting Stock Markets with Neural Networks'. Available at: https://www.duo.uio.no/handle/10852/44765 (Accessed: 18 August 2019).
21. Jia, H. (2016) 'Investigation Into The Effectiveness Of Long Short Term Memory Networks For Stock Price Prediction', *arXiv:1603.07893 [cs]*. Available at: http://arxiv.org/abs/1603.07893 (Accessed: 18 August 2019).
22. Kugiumtzis, D. and Vlachos, I. (2008) 'Turning Point Prediction of Oscillating Time Series using Local Dynamic Regression Models', *arXiv:0809.2229 [nlin]*. Available at: http://arxiv.org/abs/0809.2229(Accessed: 18 August 2019).
23. Kugiumtzis, D. (2008) 'Local prediction of turning points of oscillating time series', *Physical Review E*, 78(3), p. 036206. doi: 10.1103/PhysRevE.78.036206.
24. Majumder, M. M. R., Hossain, M. I. and Hasan, M. K. (2019) 'Indices prediction of Bangladeshi stock by using time series forecasting and performance analysis', in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–5. doi: 10.1109/ECACE.2019.8679480.
25. Sutton, R. S. (1988) 'Learning to predict by the methods of temporal differences', *Machine Learning*, 3(1), pp. 9–44. doi: 10.1007/BF00115009.
26. Werbos, P. J. (1987) 'Building and Understanding Adaptive Systems: A

Statistical/Numerical Approach to Factory Automation and Brain Research', *IEEE Transactions on Systems, Man, and Cybernetics*, 17(1), pp. 7–20. doi: 10.1109/TSMC.1987.289329.
27. R. Fang, S. Pouyanfar, Y. Yang, S.-C. Chen and S. Iyengar, (2016) Computational Health Informatics in the Big Data Age: A Survey, ACM Comput. Surv., vol. 49, no. 1, pp. 1–36
28. https://blog.algorithmia.com/introduction-to-time-series/
29. *Long Short-Term Memory | Neural Computation | MIT Press Journals* (no date). Available at: https://www.mitpressjournals.org/doi/abs/10.1162/neco.1997.9.8.1735(Accessed: 21 August 2019).
30. S.M. Kendall and K. Ord, Time Series, Oxford University Press, New York, 1997.