| S | |
|---|---|
| E1 | |
| E2 | |
| **For Office Use Only** | |

# Masters Project Final Report

# (MCS)

# 2019

| | |
|---|---|
| **Project Title** | Prediction model for predict the share market price |
| **Student Name** | Indika Anuradha Abeygunawardana Weerasinghe |
| **Registration No. & Index No.** | 2017/MCS/086<br>17440862 |
| **Supervisor's Name** | Dr. Manjusri Wickramasinghe |

| **For Office Use ONLY** |
|---|
| |
| |

# Prediction model for predict the Share Market Price

## A dissertation submitted for the Degree of Master of Computer Science

**I.A.A. Weerasinghe**
**University of Colombo School of Computing**
**2020**

## Declaration

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name: I A A Weerasinghe

Registration Number: 2017/MCS/086

Index Number: 17440862

______________________

Signature: Date:

This is to certify that this thesis is based on the work of

Mr. Indika Anuradha Abeygunawardana Weerasinghe

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:

Supervisor Name: Dr. Manjusri Wickramasinghe

______________________

Signature: Date:

## Abstract

In current days, stock market has become one of the most popular investing methods among investors. But investing on share market is not much easy thing. Most of the traders have no idea about the right time to invest in stock market. Since share market can change due to many facts, it is very hard to make the prediction about financial market trends. successful prediction can make huge profits.

Nowadays most of the investors are trying to make the predictions by analyzing the past values of share market. These predictions can be made by using machine learning techniques. Artificial Neural Networks and Support Vector Machines are the most common methods that can be used to make the predictions.

In this study, three different forecasting techniques have been implemented to make the prediction. Standard averaging, Exponential moving average and Long short-term memory are the techniques used in this article. This article is comparing the outcomes of above-mentioned methods to find out the most suitable prediction technique among them.

Here, historic data in Colombo stock exchange has been used as the inputs for above mentioned methods. The data set used in this research contains data from 2011 to 2018. Here, data preprocessing methods have to be followed as row data set which is given by CSE had some dirty, missing data. This study also talked about the preprocessing techniques that can be used to clean the dirty data.

## Acknowledgements

I would like to thank University of Colombo School of Computing, for providing a wonderful opportunity to enhance my skills and abilities.

This project became this much success because of your guidance on my research project supervisor, Dr. Manjusri Wickramasinghe, and my co-supervisor, Dr. Thilina Halloluwa. I would like to sincerely thank them for the valuable guidance, understanding, and trust that they have on me along with the support given to me in numerous ways to make this project success.

Moreover, I would like to thank Mr. Yohan Jayarathne who is working as a team lead at 99X Technology for supporting me to complete my research. I am particularly thankful for the support given by the entire team at 99X Technology company.

# Contents

# List of Figures

# List of Equations

# List of Tables

Chapter 1

# Introduction

## 1.1 Problem Domain

### 1.1.1 Stock Exchange and its behavior

Stock exchange is used as an interface between buyers and sellers. it allows investors to invest their money on stock also called as shares. Stocks represent the shares of ownership in a business and it allows traders to own a share of a public corporation.

Stock price or share price is varying time to time and it depends on the company's income. Stock price goes up If the company's economy is good or everyone thinks the company is going to do well. Most of the companies provide dividend payment to the stockholders annually [1].

Trading in stock exchange means the getting the ownership of asset of a public corporation from a seller to a buyer. In stock market each buyer has a seller. Similarly, for every seller, there is a seller. If a seller sells his shares, someone has to buy them. If there are more buyers than sellers, demand goes up and the stock price will go up. The stock price will fall if there are more suppliers than buyers [2].

There are two main ways for invest in stock market. Buy stock online and Using Full service Brokers are the main ways of investing [1].

### 1.1.2 Forecasting Stock Price

Value of an organization often determines by its share prices. Investors are always trying to forecast the future value of an organization stock as they could sell or buy shares in right time. This process is called as forecasting stock prices. Anyone can gain high profits if they cloud build high quality forecasting model to predict the shares. There is a financial theory called efficient-market hypothesis says current share market price is a reflection of all the public available information about the organization. According to EMH forecasting share value is impossible and only way to achieve high profits is jump into riskier investments. Those who are not believe the EMH are trying to forecast the share values by using various methods and technologies.

## 1.2 The problem

Stock Exchange is the place where securities such as shares, debentures issued by companies are traded. Colombo Stock Exchange (CSE) is one such market. A stock exchange provides the opportunity to the companies to raise capital, in exchange for giving investors a slice of ownership in the company. There are several individual investors, financial institutions, and companies that currently hold investment portfolios among the companies listed on the CSE.

Since the stock exchange is dynamic, there is a risk in investing money on shares. The most significant risk is that the return is not guaranteed. If a company falls or goes bankrupt, first the bondholders are getting paid. After that the shareholders will get their payments. But there is no guarantee of getting payed after a bankrupt. Though the stocks have performed well over the long term, there is no guarantee that an investor will make money on a share at any given point of time. Since the investors are quite resistant to take a risk, they always want to know the expectation of return on investment in the stock market before investing money in it. Investors in the stock market want to maximize their profits by buying or selling their investments at an appropriate time. Analyzing the historical records of a given symbol and forecasting the market price is a challenging task.

Research Question: - **How to implement a model to increase the accuracy of a stock market price predicting system by analyzing the historical data?**

this question can be answered by answering the following sub questions.

- What are the most relevant attributes that need to be used in prediction model?
- What are the ways of implementing stock market prediction system?
- What is the best neural network, that can be used to predict the stock market future trends?
- How to improve the accuracy of the created model?

## 1.3 Motivation

As we discuss above, there is a risk in investing money on shares. Predicting stock price has become an important problem in these days. If we can build a successful model for stock prediction, it will be very helpful to gain insight about market behavior over time, spotting trends that would otherwise not have been noticed. Machine Learning is the best way to address this problem by creating a prediction model.

In this research, a model has been implemented which can predict the future price of shares for some selected companied listed in Colombo Stock Exchange using historical data. Also, this research can be used to gain an idea about the different prediction methodologies.

## 1.4 Objectives

The main aim of this project is to, come up with a model to predict the stock market prices for a given company. The aim of this study has been achieved by focusing on the following objectives.

- Developing a thorough understanding of the current stock market trends.
- Performing a time series analysis for share prices for the companies listed in CSE.
- Choose the appropriate architecture and algorithm, and build the prediction model.
- Testing the predicted values' accuracy.
- Building a user-friendly tool to feed data and predict share prices.

## 1.5 Research Contribution

The research contribution of this project is to Develop a model which can be used to predict the share prices. This study is comparing three machine learning techniques and how they can be used to implement a model to forecast the share market price. This research is beneficial for Machine Learning researches and people who are willing to invest on Colombo Stock Market. In this study the prediction has been made by using exponential moving average, standard average and long short-term memory.

## 1.6 Scope

The scope of this project is to creating a solution to the research question mentioned above. This project aims to provide a tool for users to analyze the historical data of the companies listed in CSE. Scope of this project includes,

- Implementation of the models and documentation of how the forecasting models are build using machine learning techniques.
- Comparing more stock market predicting techniques is an additional outcome of this research.
- Improving the performance and the accuracy of the predicted values by evaluating the prediction model.
- Displaying Historical data and predicted values in a more readable manner (graphs and tables has been used to display the data).

## 1.7 Structure of the chapters

This section is discussing about the structure of the chapters. The first chapter provides brief introduction about the problem domain and what the forecasting is. Also, it talks about the problem, motivation, objectives, research contribution and the scope of this research.

Second chapter is the Literature review which is used to discuss about the similar researches and varieties of models used to forecast the share market prices. This chapter can also be used to get an understanding about the research gap.

Next, the methodology chapter describes the steps followed in this research to implement the proposed prediction model. Apart form the steps, this chapter also described about the problems encountered during the implementation process and how to avoid them.

Fourth chapter is all about the evaluation process and the methodologies followed to evaluate the proposed forecasting model. This chapter talks about the accuracy of the result generated by different machine learning methodologies.

Finally, the last chapter talks about the conclusion of this study and the future works related to this study.

Chapter 2

# Literature Review

## 2.1 Related Work

**Artificial Neural Network (ANN)**

ANN are analogues to nonparametric, nonlinear regression models. Therefore, ANN certainly has the potential to distinguish unknown and hidden patterns in data which can be very effective for share price prediction. This approach is suitable because there is no need of understanding the solution.

An ANN, also named as an Artificial Neural Network, is a collection of links. These links are connected to each other and they have their own weights. The idea of ANN is similar to neural networks in biology. ANN can be used to make efficient and high-quality prediction model which can be useful to achieve high profits in the field of stock marketing. Often an ANN contains three interconnected layers. First layer is input layer, which is used to fetch inputs to neural network. Then the weighted outputs of input layer will be passed to the hidden layer. Outputs of input layer will be the inputs of hidden layer. Accuracy of the process can be increased by adding extra hidden neurons. But increasing the number of neuron layers will increase the complexity of the training algorithm. Best practice is to use necessary hidden neurons rather than using larger number of unwanted hidden neurons. But this is not motivated to use a smaller number of hidden neurons. Finally, the output of hidden layer will be passed to the Output layer. ANN can be trained to perform a specific task by adjusting the weights. The weights are continually adjusted by comparing the output of the network with the target until the output of the network matches the target in the sense that the error function measuring the difference between the target and the output is minimized [3].

below Figure 1 shows the layers of neural network;

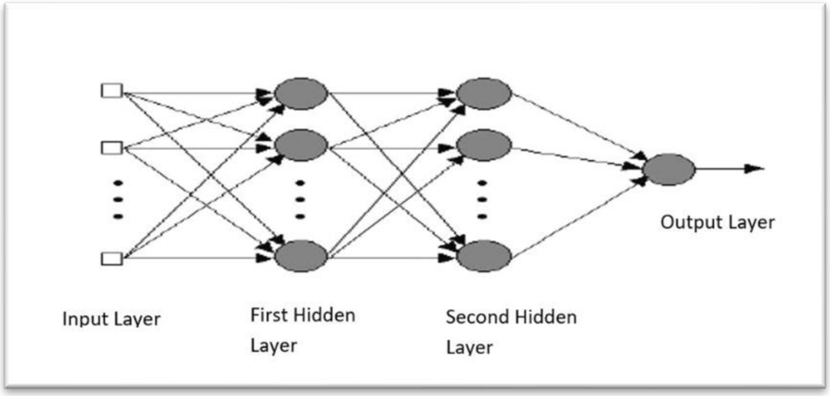


*Figure 1: Illustration of a Neural Network*

**Share value forecasting by Neural Network**

Mayan Kumar and Sunil used Feedforward Multi-Layer Perceptron (MLP) neural network to predict share prices of the companies listed under LIX15 index of NSE for the duration of 3 years. Study the current market trend and collect trend data, Build Prediction Model for the companies listed under LIX15 using MLP Neural Network Techniques and validate the model with real data are the main objectives of their research. They first co They first collected the past data of stock market and developed An MLP neural network algorithm to predict future stock price. Neural network was built and trained for different combinations of data and parameters, which are; number of input neurons, output neurons, hidden layers, neurons in each layer, learning rate and epochs using neural network toolbox. The best fit model was selected comparing the mean square errors of each model and concluded that MLP neural network gives a satisfactory output for stock price prediction. According to their research they said that Stock market data are highly time-variant and they are in a nonlinear pattern and Artificial Neural Network technique is very useful in predicting stock price of particular company [4].

According to Rosalina and Jayanto, traditional forecasting models like autoregressive integrated moving average (ARIMA) are not suitable to model financial series. They stated Artificial Neural network can be used effectively and its accuracy is high comparing traditional forecasting models. They realized that the selection of attributes while data preprocessing step, is directly affect the outcome of the model. They used Discrete Wavelet Transform Back

Propagation Neural Network (DWT-BPNN) and Maximum Overlap Wavelet Transform Back Propagation Neural Network (MODWT-BPNN) on the data collected from Indonesian stock market (Code-JKSE) for the duration five years (from January 2010 until March 2015). The results stated that MODWT-BPNN performed well on the dataset. According to them Wavelet is a wave with amplitude begins at zero, increases and then decreases back to zero and it can be used as a powerful signal processing tool as it is built to have specific attributes. For the testing they used three methods. Full training data set, which uses all the data set used in training. In this scenario testing dataset is a part of the training dataset. Second method is Half Training and new data set. In second scenario half of the dataset is taken from the training dataset and other half is completely new dataset. Third scenario is Full new dataset which is using completely new dataset for testing. According to Rosalina and Jayanto, MODWT-BPNN performed best when the testing dataset is completely new. So, they recommending to use the third test method [5].

According to Zhang, Fulin, Bing, Wenyu, Qiongya and Ting the common methods forecasting models such as; Box-Jenkins model, Autoregressive (AR) model, Autoregressive Integrated Moving Average (ARIMA) model assume that the time series needs to be linear and stationary, hence the mentioned models are not suitable to forecast stock prices as stock prices are non-linear and dynamic. They have used Levenberg-Marquardt backpropagation (LM-BP) to predict the opening price, highest price, lowest price and closing price and stated that the predicted values are accurate [6].

According to Soni's survey, Artificial Neural Network can be used very effectively in prediction problems as it has some special characteristics. First, even if the relationship of the output and input are very complicated, artificial neural network can find the relationship between them. It means artificial neural network can be applied to extracting the relationships among data even the relationship is very hard to find. Second, artificial neural network can recognize new patterns using completely new training data set after training. It means artificial neural networks have generalization ability. Artificial neural network can be used in most of the pattern recognition systems, since they predicting unseen data using training data. Third neural network helps to understand any complicated relationship between the input and the output of the system and it has been called as general function approximations. As mentioned below she stated that there are three strategies to do the financial prediction in artificial neural networks,

- Artificial neural network can use its inputs to find the rules associated with the current state of the system which can be used to make the predictions [7].
- It can be used to find the relations between past and future conditions [7].
- When designing the ANN, it can be designed to accept larger number of inputs (past data) to find out the connection between past and future trends. To find the connections, recurrent connections have to be used [7].

in 2017, Samarawickrama and Fernando conducted a study to develop model based on Recurrent Neural Network (RNN) to predict stock prices of the selected companies listed on Colombo Stock Exchange (CSE). They built models based on Feedforward, Simple Recurrent (SRNN), Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) networks. The study was conducted to predict high price and low price and concluded that the best fit is the feedforward network, as it returned lowest forecasting errors and the maximum percentage accuracy [8].

Senanayake and Jananathan have developed a stock market prediction system using Artificial Neural Network and they collected Data from Colombo Stock Market Exchange. They observed the mean square error on the network by changing the network architecture for each training to find the optimal number of neurons in hidden layer. They chose (6-6-2) architecture (6 input neurons, 6 hidden layer neurons and 2 output layer neurons) with learning rate of 0.2. They stated that the most common network architecture is the backpropagation method as it offers good generalization abilities and it’s relatively straightforward to implement. The performance of this backpropagation network can be improved by using recurrence or reusing past inputs and outputs [9].

In 2018, Tang and Chen created a hybrid model combining Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) networks for time series data and Convolution Neural Network (CNN) for abstract high dimensional data. According to results, the hybrid model forecasted stock market future prices to a great degree of accuracy and the it outperformed the baseline [10].

Song, Zhou and Han have conducted a study to predict share prices of selected three companies listed in Shanghai Stock Exchange. Prices were predicted using five neural network models namely; backpropagation, radial basis function, general regression, support vector machine and least squares support vector machine networks. The conclusion of the study was that the back propagation neural network outperformed other models consistently and robustly [11].

Vrbka and Rowland carried out the study in 2017, to find the best Multi-Layer Perceptron (MLP) network to predict the share prices of the selected companies listed in Prague Stock Exchange. According to the study, the networks with two layers and four layers predicted the prices with a great accuracy [12].

**Exponential moving average**

There are several moving averaging methods that can be used to make predictions. An exponential moving average (EMA) method is one of the most used averaging method. It is also called as exponentially weighted moving average (EWMA).

Moving averages is the average price of an instrument over a specific period of time. Moving average methods can be categorized into several types. This categorization can be done based on the way of assigning weights to the data points.

The exponential moving average is a line on the price chart based on a mathematical formula to smooth out the price. Placing more weight on the recent price and less weight to past prices, the EMA adapts more quickly to the latest price changes in price. Below Eq. 1 shows the equation to calculate the EMA. Here close is the closing price and n is the number of data points. [13].

$$Exponential\ moving\ average = (close - Previous\ EMA) \times \left(\frac{2}{n+1}\right) + \ Previous\ EMA$$

*Equation 1: Equation to calculate EMA*

In 1992, Robert Edwards and John Magee have carried out a research about stock trends. During their research they realized that the moving average has several categories. Simple Moving average, Weighted or Exponential moving average and linear moving average are the categories that they could find. Outcome of their research says, that SMA can be used to make prediction using computers. But working with exponential moving averaging is easy when compared to others. It can also use with a smaller number of past facts. Before calculate the future values of the stocks using averaging methods, it is must to select correct smoothing constant [14].

During the 2007 and 2008 periods there was an economic crisis. After that, in 2009 Girdzijauskas et al. conducted a research to find out the future market bubbles and the ways of facing them. They realized that the best way to forecast the processes is to use the exponential models [15].

According to Ehlers, Exponential moving average (EMA) is considered to be a better tool than a simple moving average. Because it attaches greater weight to current data and changes in price correspond faster than with the simple one [16].

James Jr had the same opinion as Ehlers. He said that the exponential moving average helps to predict the future price movement and use it as an aid to minimize the losses. But later in his studies, he found that exponential moving average is not an effective tool in predicting the future price movement [17].

Three researches, Josip Arneric, Elza Jurun and Snježana Pivac have stated that it is must to have technical and fundamental analysis to forecast the share values. Technical analysis helps to discover the price movements of shares. Then the fundamental analysis has to be done to forecast the share values by looking at the company fundamentals. Here above-mentioned researchers have focused on only the technical analysis. They stated that the trend can be either time structure or general direction. In this research Josip, Elza and Snježana have used exponential moving averaging method to conduct their research as the simple moving average method has drawbacks of finding long and short-term strategies. It is difficult to find reliable values by using simple averaging method to buy and sell shares [18].

In 1999, Murphy has said that the validity of the moving average method can be verified easily and the accuracy of the prediction made by using moving average is very explicit. This is the main reason to enhance the popularity and the usage of the moving average. He also said that the moving average is the base for several timeseries analyzing technologies.

According to the way of calculating the averages, moving average can be implemented in various ways. As an example, to calculate the future value of the shares, exponential moving and linear moving averages can be used instead of the simple moving average. Here the exponential averaging method uses exponential average and linear moving average uses linear average to calculate the averages. Without using single averaging method, it is ok to use two averaging methods together to calculate the future share values. The prediction can be found when two averages cross each other. Here the two averaging methods use two different time gaps to calculate the prediction [19].

Other than the linear and the exponential moving averaging methods, simple moving average is the most commonly used averaging mechanism to calculate the averages. But there are few drawbacks of using the simple averaging method. One popular drawback is that the simple moving average uses only the price information in a given time period. It doesn't consider the other useful information. This will be an issue of generating accurate prediction as it neglecting the valuable information. Another drawback is that the simple averaging is assigning same weight for all the prices. It is proven by researches that bigger weights have to be assigned for recent values to generate an accurate output. Exponential and linear moving average methods are solutions to avoid these issues [19].

As mentioned before, moving averages methods have many benefits and they have been widely used to calculate the future values when compared to simple chart analyzing technologies. But they cannot be applied for each and every situation as they have some major issues. These methods are working fine only if the stock market has high or low trends as they use only the price trend. Moving average is not working properly if the rate of changing the price of shares is high, or the share price variance is not even. So, the investors who are trying to forecast the share value, are informed not to depend on the moving averaging methods to forecast the share price if there are no clear stock trends [19].

**Share price prediction using Time Series Analysis**

Devi, Sundar and Alli carried out the study with the objective of building the forecasting model which gives the minimum forecasting error. The study used ARIMA model with different parameters for top 4 companies listed in the NSE-Nifty Midcap 50 index and derived a best model equation [20].

Two researchers, Angadi and Kulkarni developed a prediction model in 2005 to forecast stock market by using the ARIMA model and concluded that this model is fitting reasonably well in predicting stock market [21].

In 2014 Adebiyi, Adewumi and Ayo built a share price predictive model using the ARIMA model. The closing price of the listed companies in New York Stock Exchange and Nigeria Stock Exchange were used in building the model. The best fit ARIMA model demonstrated that the model is the most suitable to predict share prices satisfactorily on a short-term basis [22].

Kamalakannan, Indrani and Sneha used the ARIMA model for the historical closing price data of Apple stock as this model has been vastly used in economics and finance fields. The study concluded that ARIMA is a better model in predicting the short-term trends of the stock market [23].

Ashik and Kannan conducted a study in predicting Nifty 50 index in National Stock Exchange India. They evaluated and predicted the trend and fluctuations of the index through Box-Jenkins methodology. The study concluded that the fitted model has great accuracy for prediction as it returned a higher R-square value and a lower Mean Absolute Percentage Error (MAPE) value [24].

Ding, Fang and Zuo carried out a study to predict the stock market movement by analyzing the behavior of S&P 500 index. They built three models for this prediction based on Support Vector Machine (SVM), logistics regression and neural networks and concluded that SVM was the most accurate model for the prediction, but the statistical tests resulted that SVM is not significantly better than logistic regression [25].

In 2017, Jadhav and Kamble studied the trends of closing prices of selected four oil companies in which were listed in the Bombay Stock Exchange (BSE), using ARMA model. Based on the fitted models they predicted the closing prices of each company [26].

An Indian researcher called Kuttichra used ARIMA model to predict share prices of selected two pharma companies listed in NSE India in 2017. According to the study ARIMA model is said to be an efficient model to analyze the trends and patterns of a share price time series, but it is not efficient for long term predictions [27].

## 2.2 Conclusion

Here after the literature review we can clearly see that predicting share market price is an important issue in finance. This was a motivational factor for investors and brokers to invest money in the share market. Most of the researchers has used Neural Networks and Time Series analysis methods like ARIMA. Here we can see, Neural Network techniques can produce better and high-performance output than Time Series analysis techniques. Neural Network technique called Long Short-Term Memory is going to be used in this research.

Chapter 3

# Methodology

## 3.1 Introduction

This section describes the way of addressing the above-mentioned research question. It contains below steps.

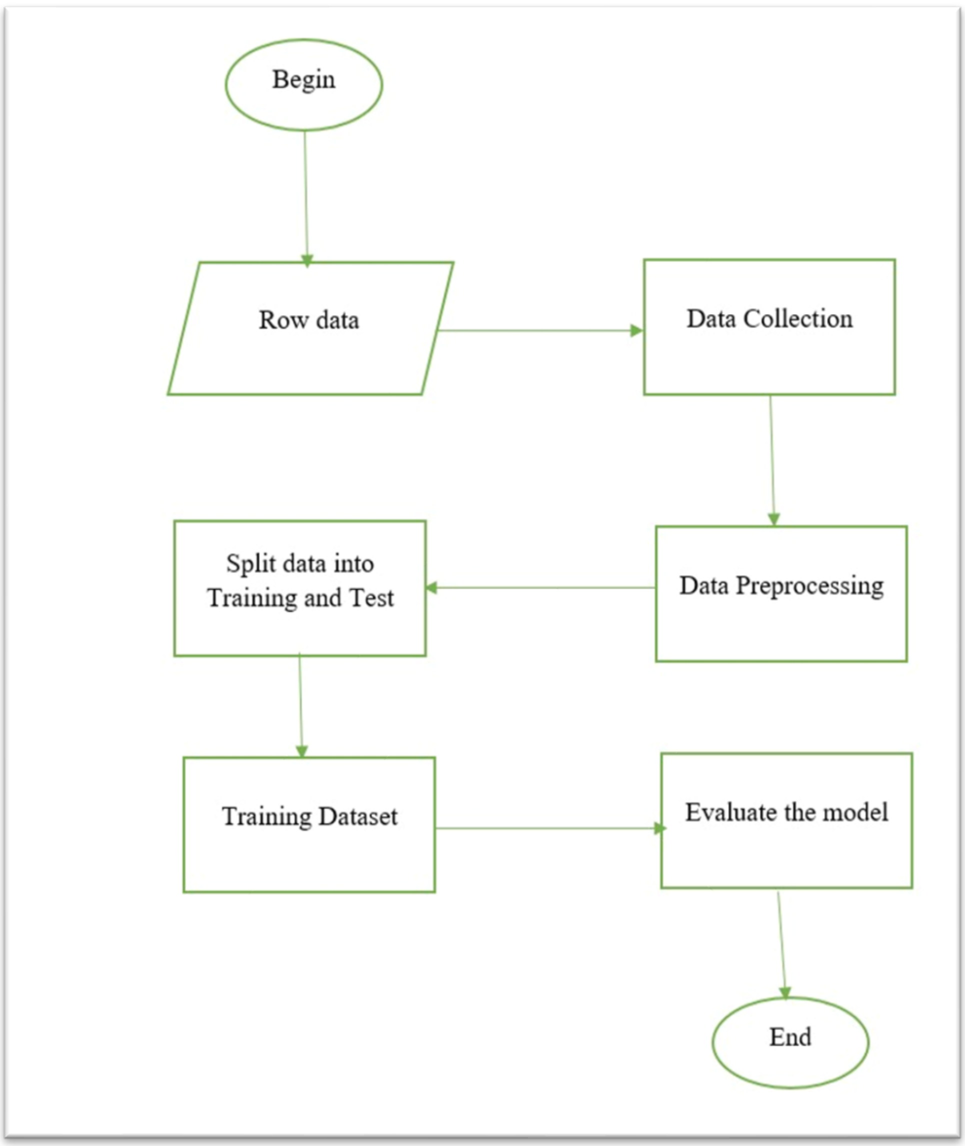


*Figure 2: Steps need to be followed to make the prediction*

## 3.1.1 Data Collection

Past data of the stock market has been collected from the Colombo Stock Exchange. CSE provides data as separate Excel files. It contains data from 2011 to 2018. The data set contains below attributes.

- Company Id.
- Company name.
- Day: - data collected date.
- Date High: - highest price recorded date.
- High: - highest price for the day (Rs).
- Date Low: - lowest price recorded date.
- Low: - Lowest price for the day (Rs).
- No of Trades.
- No of shares (stocks).
- Turnover.
- Last Traded.
- Days Traded.

| DAILY HIGH LOW AND CLOSING PRICE OF 02ND JANUARY TO 31ST DECEMBER 2012 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Company Id | : AAF | Security | Type : N | Sub Type | : 0000 | | | | | |
| Short Name | : ASIA ASSET | | | | | | | | | |
| Day | Date High | High (Rs.) | Date Low | Low (Rs.) | Closing (Rs.) | Trades(No.) | Shares(No.) | Turnover(Rs.) | Last Traded | Days Traded |
| 1/12/2012 | 12-Jan-12 | 4.30 | 12-Jan-12 | 3.00 | 4.20 | 2,428 | 38,538,200 | 144,153,480.00 | 12-Jan-12 | 1 |
| 1/13/2012 | 13-Jan-12 | 6.30 | 13-Jan-12 | 4.50 | 5.80 | 3,132 | 40,938,200 | 226,636,450.00 | 13-Jan-12 | 1 |
| 1/17/2012 | 17-Jan-12 | 8.70 | 17-Jan-12 | 6.30 | 8.50 | 3,625 | 45,760,300 | 365,896,240.00 | 17-Jan-12 | 1 |
| 1/18/2012 | 18-Jan-12 | 8.90 | 18-Jan-12 | 5.20 | 5.50 | 2,628 | 21,117,600 | 152,660,190.00 | 18-Jan-12 | 1 |
| 1/19/2012 | 19-Jan-12 | 7.50 | 19-Jan-12 | 4.70 | 7.10 | 3,231 | 23,529,800 | 156,061,890.00 | 19-Jan-12 | 1 |
| 1/20/2012 | 20-Jan-12 | 8.70 | 20-Jan-12 | 6.00 | 6.20 | 1,955 | 23,947,200 | 159,884,800.00 | 20-Jan-12 | 1 |
| 1/23/2012 | 23-Jan-12 | 6.40 | 23-Jan-12 | 5.60 | 5.90 | 867 | 5,991,400 | 35,911,810.00 | 23-Jan-12 | 1 |
| 1/24/2012 | 24-Jan-12 | 6.20 | 24-Jan-12 | 5.60 | 6.10 | 808 | 7,611,400 | 46,157,340.00 | 24-Jan-12 | 1 |
| 1/25/2012 | 25-Jan-12 | 6.70 | 25-Jan-12 | 6.10 | 6.50 | 860 | 8,396,900 | 53,833,420.00 | 25-Jan-12 | 1 |
| 1/26/2012 | 26-Jan-12 | 7.10 | 26-Jan-12 | 6.70 | 7.10 | 1,040 | 11,050,700 | 77,488,520.00 | 26-Jan-12 | 1 |
| 1/27/2012 | 27-Jan-12 | 7.50 | 27-Jan-12 | 6.80 | 7.10 | 1,414 | 14,803,100 | 106,549,120.00 | 27-Jan-12 | 1 |
| 1/30/2012 | 30-Jan-12 | 7.10 | 30-Jan-12 | 6.40 | 6.70 | 476 | 2,977,900 | 20,158,930.00 | 30-Jan-12 | 1 |
| 1/31/2012 | 31-Jan-12 | 6.90 | 31-Jan-12 | 6.50 | 6.60 | 314 | 1,564,800 | 10,332,830.00 | 31-Jan-12 | 1 |
| 2/1/2012 | 1-Feb-12 | 6.80 | 1-Feb-12 | 6.10 | 6.20 | 304 | 1,825,100 | 11,416,690.00 | 1-Feb-12 | 1 |

*Figure 3: Sample of row data in excel*

Turnover can be used to measure the liquidity of stock. Stock Liquidity shows how ease of converting stocks into cash. High Liquidity stocks can be quickly sold or bought in the market. Turnover can be calculated by dividing the total number of shares traded over a given time range by the average number of shares outstanding for the period.

$$Turnover = \frac{number\ of\ shares\ traded\ over\ a\ period}{average\ number\ of\ shares\ outstanding\ for\ the\ period}$$

*Equation 2: Equation for calculate the turnover*

### 3.1.2 Data Preprocessing

Real world data is often not complete, in-consists and contains errors. Preprocessing is a technique of resolving these issues. It is a data-mining technique that transforms raw data into an understandable format. Python has libraries that can be used to preprocess the data. Three common libraries are Numpy, Matplotlib and Pandas. Panda can be used to read and write data into files and Numpy can be used to preprocessing the data set. As the first step, data has to be loaded to a preprocess library to remove error-prone data. the same library can be used to splitting data set into training and test set. The training set can be used to train the model, and the test set can be used to evaluate the proposed model.

Before loading the data to Numpy, all the data has to be formatted into a single file. Data sets are in different files and the way of saving them in files is difficult to read by a program.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 39 | 12/31/2012 | 31-Dec-12 | 3.00 | 31-Dec-12 | 2.60 | 2.90 | 186 | 1,707,498 | 4,813,743.70 | 31-Dec-12 | 1 |
| 40 | | | | | | | | | | | |
| 41 | Company Id | : AAIC | Security | Type : N | Sub Type | : 0000 | | | | | |
| 42 | Short Name | : ASIAN ALLIANCE | | | | | | | | | |
| 43 | Day | Date High | High (Rs.) | Date Low | Low (Rs.) | Closing (Rs.) | Trades(No.) | Shares(No.) | Turnover(Rs.) | Last Traded | Days Traded |
| 44 | 1/2/2012 | 2-Jan-12 | 181.00 | 2-Jan-12 | 165.10 | 165.20 | 17 | 9,700 | 1,642,670.00 | 2-Jan-12 | 1 |
| 45 | 1/3/2012 | 3-Jan-12 | 169.00 | 3-Jan-12 | 169.00 | 169.00 | 1 | 200 | 33.800.00 | 3-Jan-12 | 1 |

*Figure 4: Sample of row dirty data in excel*

Unnecessary texts like “Company ID”, “Security Type”, “Sub Type” and “Short Name” have been removed from the Excel using the Excel commands.  For the prediction model, company ID is a necessary attribute. It should be added to a new column. The column has been named as Company ID.

| A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ) | Day | Date High | High | Date Low | Low | Closing | Trades(No.) | Shares(No.) | Turnover(Rs.) | Last Traded | Days Tr | Company ID |
| 1 | 1/2/2013 | 2-Jan-13 | 3.20 | 2-Jan-13 | 2.90 | 3.00 | 467 | 4,165,602 | 12,628,887.40 | 2-Jan-13 | 1 | AAF |
| 2 | 1/3/2013 | 3-Jan-13 | 3.20 | 3-Jan-13 | 2.90 | 3.00 | 194 | 2,510,471 | 7,567,663.10 | 3-Jan-13 | 1 | AAF |
| 3 | 1/4/2013 | 4-Jan-13 | 3.10 | 4-Jan-13 | 2.90 | 3.00 | 66 | 245,201 | 735,505.40 | 4-Jan-13 | 1 | AAF |
| 4 | 1/7/2013 | 7-Jan-13 | 3.00 | 7-Jan-13 | 2.90 | 2.90 | 50 | 352,628 | 1,027,472.10 | 7-Jan-13 | 1 | AAF |
| 5 | 1/8/2013 | 8-Jan-13 | 3.00 | 8-Jan-13 | 2.80 | 2.90 | 78 | 66,156 | 191,161.20 | 8-Jan-13 | 1 | AAF |
| 6 | 1/9/2013 | 9-Jan-13 | 2.90 | 9-Jan-13 | 2.80 | 2.80 | 31 | 78,353 | 222,853.70 | 9-Jan-13 | 1 | AAF |

*Figure 5: Sample of preprocessed data in excel*

Once all the data has been changed in to a single format, then they can be merged into a single file.

After all the excel manipulations, Data can be loaded into the preprocessing libraries. Using Pandas library data can be loaded. Pandas has various methods to input/output manipulations. Even it can Read and Write data to SQL Query or Database Tables. Since data is read from an Excel file, the Pandas **read_excel** method can be used to load the data into a Pandas Data Frame.

```
read_excel (io [, sheet name, header, names, …])
```

*Figure 6: Numpy method that can be used to read excel data*

Sometimes, especially when reading data from files which have values as numbers, some of them can be read as string, or vice versa. For fix this issue, Data types have to be Normalized as below.

```
Pandas.read_Excel(‘CSEDataSet.csv’, dtype={‘High’: int})
```

*Figure 7: Numpy method that can be used to normalized the data*

After the normalization, High column value will always read as numeric. Once the data is loaded to the DataFrame, then it can be clean using Numpy and Pandas Python libraries.

In this data set, there were some fields with missing values as shown in below Figure 8.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 21-Nov-13 | 4.10 | 21-Nov-13 | 4.10 | 4.10 | 2 | 200 | 820.00 | 21-Nov-1 |
| 22-Nov-13 | | 22-Nov-13 | | | 1 | 18,000,000 | 90,000,000.00 | 22-Nov-1 |
| 25-Nov-13 | 4.00 | 25-Nov-13 | 4.00 | 4.00 | 2 | 4,100,100 | 20,500,400.00 | 25-Nov-1 |

Figure 8: Sample of missing data in excel

As the data set contains few numbers of rows with missing values, they can be removed from the data set. To remove those fields, first those values should be replaced with Numpy's NaN values as below. "^\s*$" regex can be used to find the missing values fields.

```
dataFrame.replace(r'^\s*$', numpy.NaN, regex=True)
```

Figure 9: Numpy method to find the NaN values in data frame

A Pandas Index extends the functionality of Numpy arrays to allow for more versatile slicing and labeling. This index will be applied to the Uniquely Valued identifying field. This data set contains 'ID' column which contains unique values. Index can be applied to the ID column.

```
dataFrame[ID].is_unique
```

Figure 10: Numpy method to make unique data frame row Id

After all the above steps, below code has been used to validate the dataset. It returns all the columns and their null values count.

```
print(dataFrame.isna().sum())
```

Figure 11: Numpy method to get the NaN value count

Below Figure 12 shows the output of above code. The output contains all the attributes and their null values count as mentioned above.

```
ID                            0
Day                           0
High                          0
Low                           0
Closing                       0
No. of Trades                 0
No. of Shares                 0
Turnover                      0
CompId                        0
SecType                       0
Year                          0
Average(High & Low)           0
difference Closing/Average    0
dtype: int64
```

*Figure 12: Attributes and their NaN value counts*

### 3.1.3 Split data into Training and Test

Usually in data science, Dataset is split into training and test data. The known output is called as training data and model learns on this data. Test dataset is used to test the model's prediction.


*Figure 13: Training and test datasets*

Here if the dataset is not split well, one of two things might happen.

- Overfitting

  Overfitting occurs when the model is too complex. This model can generate very accurate data for the training data but it might not produce many accurate data on untrained or new data because this model is not generalized. This model learns the "noise" in the training data instead of the actual relationship between variables in the data. But this noise is not in any other dataset and cannot be applied to it.

- Underfitting

  This happens when the model does not fit the training data and misses the trends in the data. This occurs as the result of very simple models. For example, if we try to fit a linear model to data that not linear. These kinds of models might have poor predictive ability.

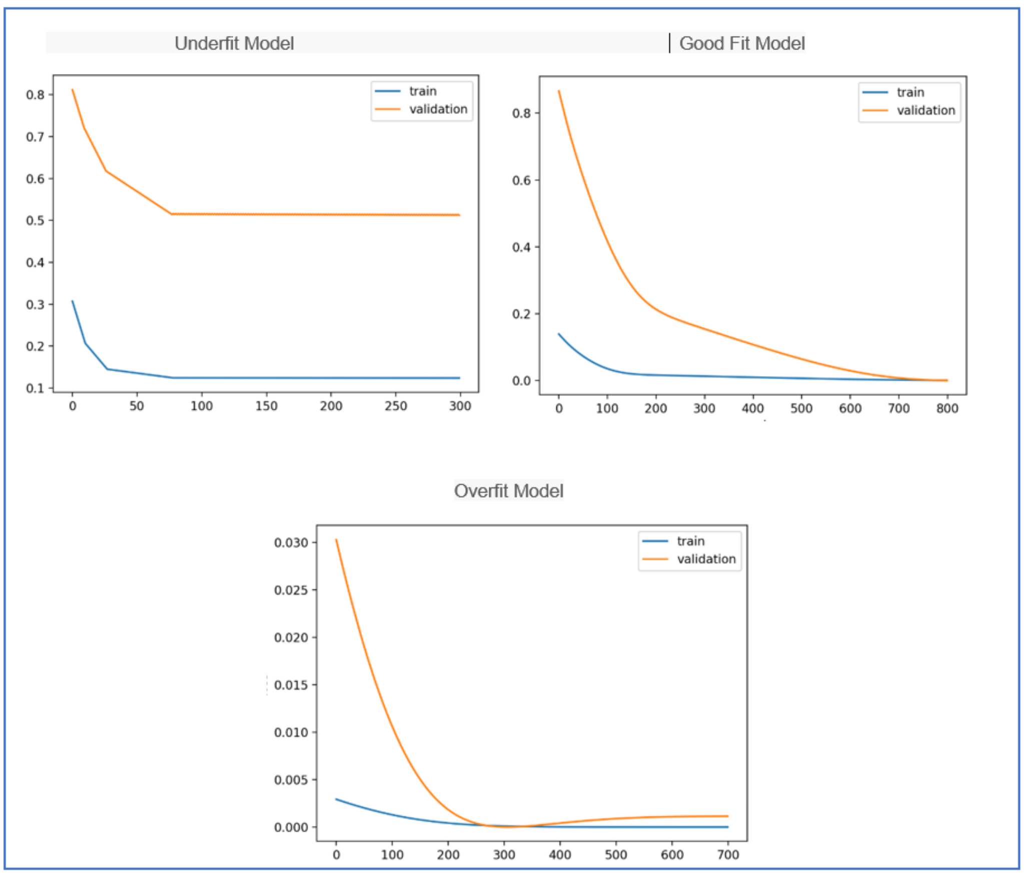


*Figure 14: Underfitting and Overfitting graphs*

Train/split and cross validation help to avoid overfitting situation. To avoid underfitting situations, we have to select best model to make the predictions [28].

To split the data set Sklearn python library can be used. Data can be divided into two parts. One part contains 80% of data and it is used as the training set. The other 20% is the test set.  As the data set contains data for eight years, the data set can be divided into two parts. Training set can have five years of data and the test set can contain three years of data. The general thumb rule is that the bigger the training data better the result. It is generally true. But it also depends on the quality of the training data. That's why the preprocessing process is required.

### 3.1.4 Training Dataset

Training dataset is a collection of sub processes.

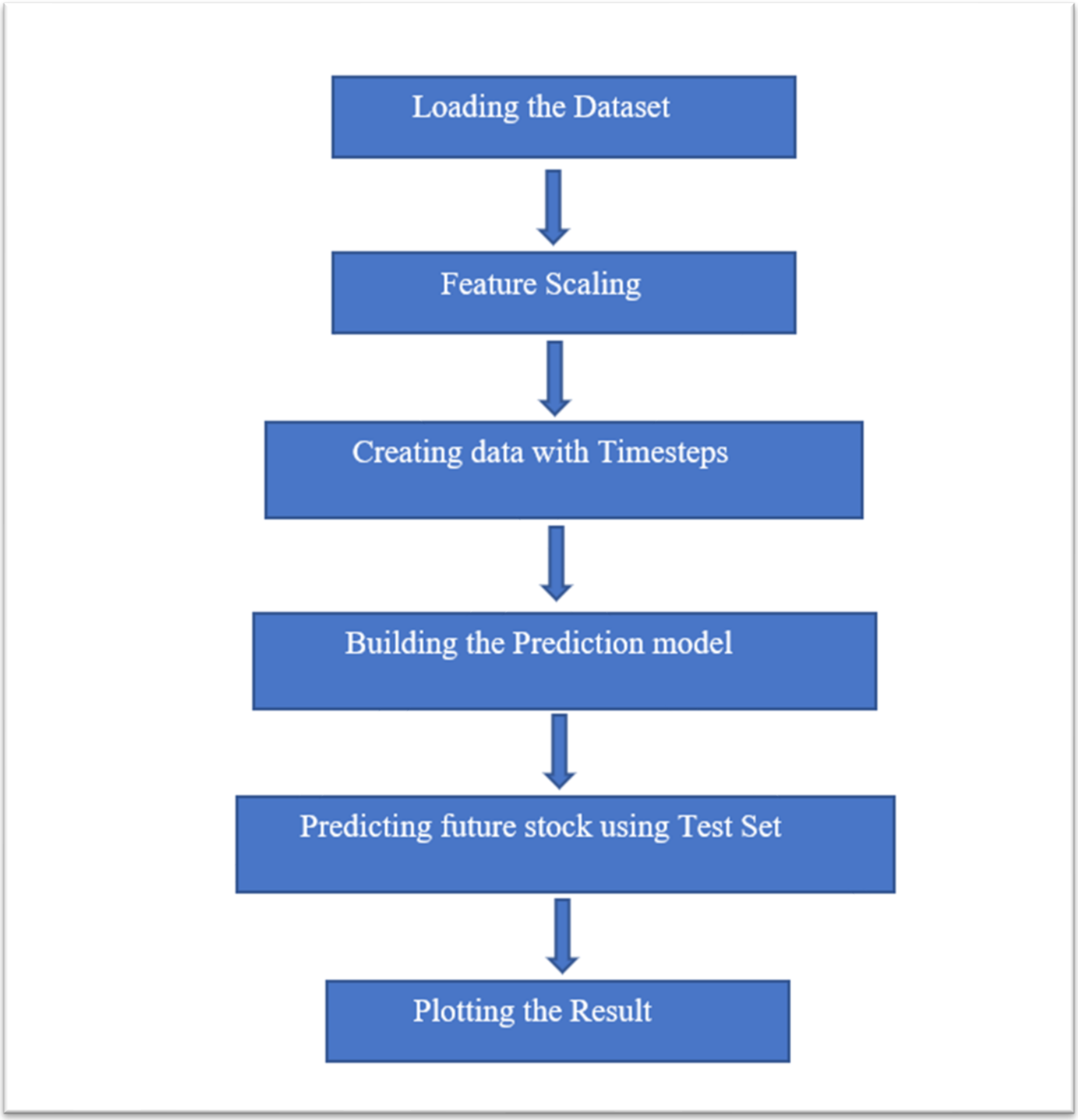


*Figure 15: Process of training the dataset*

Selecting and Developing appropriate network model will help to increase the accuracy of the output. Multilayer Perception and Long Short-Term Memory have widely used two network types. Here Long Short-Term Memory has been selected. Here TensorFlow has been used for the implementations. TensorFlow is an end-to-end open source platform that can be used in machine learning. As It has a comprehensive, flexible ecosystem of tools and libraries, developers can easily build and deploy machine learning powered applications. It also has a larger number of community resources.

Long Short-Term memories are designed to contain information for longer time periods. This has been developed to solve the vanishing Gradient Problem with Recurrent Neural Networks. Inputs for the neural network should be the lowest value, the highest value and the Turnover in the previous days. Other information in the dataset should not be concerned as the goal is to predict the stock share using the stock value share history. So, this model can be categorized as a time series prediction model. In this neural network there should be three neurons in the input layer to get the lowest value, the highest value and the Turnover value. In the hidden layer, all the neurons should be connected to the input and output layer. There is one neuron in the output layer which predicts the share price of the next day of the stock market. There can be neural networks with number of layers and hidden units. They can be used to learn a complex representation of the data. But the network computation of this kind of network is very expensive. Choosing appropriate number of neurons for the hidden layer is a challenging task. As per the Jeff Heaton, the number of neurons in hidden layer can be decided by using many rules as shown below [29].

- Number of neurons in hidden layer should be larger than the size of the input layer and it should be smaller than the size of the output layer.
- The number of hidden neurons should be 2/3 the size of the input layer, plus the size of the output layer.
- The number of neurons in hidden layer should not be larger than twice the size of the input layer.

Here once the dataset is preprocessed, first thing to do is Normalize the Data. Normalizing will help to changing the values of numeric columns to a common scale. This will increase the performance of the model. Here Scikit-Learn's MinMaxScaler has been used to scale the training dataset with numbers between zero and one.

```
from sklearn.preprocessing import MinMaxScaler

sc = MinMaxScaler(feature_range=(0, 1))
trainData = sc.fit_transform(close_prices)
print(trainData.shape)
```

*Figure 16: Implementation of MinMaxScaler*

The next step would be incorporating Timesteps into data. Here, the data has to be provided in the form of 3D array to the LSTM model. Here, this study has used data in 60 Timesteps before converting it into a Numpy array. Then the data has been converted to a 3D array with 60 Timesteps and one feature at each step. The implementation has depicted in below Figure 17.

```
X_train = []
y_train = []
n = dataList.size

for i in range(60, n):  # 60 : timestep
    X_train.append(trainData[i - 60:i, 0])
    y_train.append(trainData[i, 0])

X_train, y_train = np.array(X_train), np.array(y_train)
X_train = np.reshape(X_train, (X_train.shape[0], X_train.shape[1], 1))  # adding the batch_size axis
print(X_train.shape)
```

*Figure 17: Incorporating the Timesteps into data*

Then the next step is creating the LSTM model. In this study Keras has been used to create the LSTM model. Few Keras imports have to be made to develop the LSTM as below.

```
from keras.models import Sequential
from keras.layers import LSTM
from keras.layers import Dropout
from keras.layers import Dense
```

*Figure 18: Importing the Keras libraries*

As mentioned in above Figure 18, this study has used four Keras libraries. To develop the neural network, library called Sequential has been used. Next important library is LSTM. It can be used to build the LSTM layer. There should be a mechanism to prevent the overfitting in layers. It can be easily achieved by using Dropout library. Library called Dense has to be used to make the connections within layers.

In this study 0.2 has been specified as the Dropout layer. That means 20% of layers will be dropped. To compile the prediction model Adam optimizer has been used and set the loss as MSE. Then the model has set to run for 20 epochs with the batch size of 32 as follow.

```
model = Sequential()

model.add(LSTM(units=100, return_sequences=True, input_shape=(X_train.shape[1], 1)))
model.add(Dropout(0.2))

model.add(LSTM(units=100, return_sequences=True))
model.add(Dropout(0.2))

model.add(LSTM(units=100, return_sequences=True))
model.add(Dropout(0.2))

model.add(LSTM(units=100, return_sequences=False))
model.add(Dropout(0.2))

model.add(Dense(units=1))
model.compile(optimizer='adam', loss="mean_squared_error")

model.fit(X_train, y_train, epochs=20, batch_size=32, verbose=2)
```

*Figure 19: Initialize the NN and add LSTM layers*

Batch Size indicates number of input samples that can be seen by the prediction model before updating the weights. Assume there are 100 samples and the weights should be updated every time the NN has seen an input. To satisfy the above scenario, hundred batches have to be used and the batch size should be one. Imagine batch size has been changed to one and hundred batches have been used in a neural network. Then the prediction model can see all the inputs before it changes its weights. Efficiency of the prediction model is depending on the batch size. Larger number of batches increase the training speed and small number of batches will reduces the speed of training. But the size of the batch has a direct connection with generality of the model. Larger number of batches consume more memory and it will reduce the generality of the model. Here various values had to be tried out as the batch size to find out the right batch size for the model.

Once the model is created next step would be making prediction on the test set. The test data set has to be imported as the first step. Before predicting, test data set has to be modified as training set. Timestep can be set as 60 and using MinMaxScaler data can be reshaped. Then the created model can be used to make the prediction. The predicted prices are not in a normal readable format. To make them readable inverse_transform has to be used.

```
testData = pd.read_excel('dataNewLSTM.xlsx')
testData = testData[testData['CompId'] == 'AAIC']
testData = testData.iloc[:, 2:3]
y_test = testData.iloc[60:, 0:].values
inputClosing = testData.iloc[:, 0:].values
inputClosing_scaled = sc.transform(inputClosing)
inputClosing_scaled.shape
X_test = []
length = len(testData)
timestep = 60
for i in range(timestep, length):  # doing the same preivous preprocessing
    X_test.append(inputClosing_scaled[i - timestep:i, 0])
X_test = np.array(X_test)
X_test = np.reshape(X_test, (X_test.shape[0], X_test.shape[1], 1))
X_test.shape

y_pred = model.predict(X_test)
predicted_price = sc.inverse_transform(y_pred)
```

*Figure 20: Making prediction on test data*

After the prediction, the main part is visualizing the models output in a human readable format. This study has used the graphs to plotting the result. That can be done by using Matplotlib library. Two Line graphs have been used to visualize the actual and predict values.

```
import matplotlib.pyplot as plt

plt.plot(y_test, color='blue', label='Actual Stock Price')
plt.plot(predicted_price, color='red', label='Predicted Stock Price')
plt.title('Stock Price Prediction')
plt.xlabel('Time')
plt.ylabel('Stock Price')
plt.legend()
plt.show()
```

*Figure 21: Visualizing the prediction*

### 3.1.5 Fine Tuning the model.

Creating the model and visualizing its result is not enough for the prediction models. Once the prediction model is created, the model has to be fine-tuned. This research has used two options to fine tune the prediction model.

- Hand Tuning or Manual search.
  Here Prediction model is being tuned by using specific values for each parameter one by one. This is the most painful fine-tuning method and it might take number of rounds to find the correct combination. But the performance of this process can be improved by the experience and analyzing the initial result. Initially manual search has been used in this study. Then the below mentioned Grid Search method has been implemented.

- Grid Search.
  Grid Search method has automated the painful Manual Search method. In grid search, values for the parameters that need to be optimized is passed as a range. Then the model is trained using the

given parameter values and find the validation lost for each combination. This is a time-consuming method.

Grid search can be implemented by using SK-Learn models GridSearchCV directly. Also, Keras has their own model to implement the Grid Search. But it cannot be used directly. SK-Learn has the advantage of running jobs in parallel.

If above mentioned two libraries are not working, grid Search can be implemented as below. It will reduce the overhead of learning the syntax for a new package.

```
search_params = {
    "batch_size": [30, 31, 32],
    "time_steps": [30, 60, 90],
    "lr": [0.01, 0.001, 0.0001],
    "epochs": [20, 30, 50]
}


def get_all_combinations(params):
    all_names = params.keys()
    combinations = it.product(*(params[name] for name in all_names))
    return list(combinations)


def run_search(mat, params):
    param_combs = get_all_combinations(params)  # list of tuples
    logging.info("Total combinations to try = {}".format(len(param_combs)))
    for i, combination in enumerate(param_combs):
        logging.info("Trying combo no. {} {}".format(i, combination))
        eval_model(mat, combination, i)


run_search(X_train, search_params)
```

*Figure 22: Implementation of Grid Search*

# Chapter 4

# Evaluation

## 4.1 Overview

### 4.1.1 Research Scope

The research scope of this project is to identify a model that can be used to predict the stock market price. Forecasting stock market share price is an important financial issue that has attracted stock market investors' and researchers' attention for many years. As mentioned in the literature review researchers are trying to find the predictive relationships between the past and future stock market trends. The past information on the stock market can be found from various sources. For this research past data has been collected from the Colombo Stock Market Exchange. This study tries to analyses the different time series analysis methods to decide the better model to predict the best time for buying or selling stocks based on the knowledge extracted from the past stock market data. Following is the main research question that is going to be answered by this research.

"How to implement a model to increase the accuracy of a stock market price predicting system by analyzing the historical data?"

Also, the research problem contains below sub questions and they have been answered by the current implementation.

- What are the most relevant attributes that need to be used in prediction model?
- What are the ways of implementing stock market prediction system?
- What is the best predictive model, that can be used to predict the stock market future trends?
- How to improve the accuracy of the created model?

### 4.1.2 Current state of the problem domain

As mentioned in literature review, above mentioned research question can be addressed in a number of ways. Different studies have used different techniques and tools to predict the future stock market trends. Here this study has analyzed three time series analyzing techniques that can be used to predict the future trends in Colombo stock exchange. Followings are the techniques which are used in this study.

- Exponential Moving Average
- Standard Average
- Long Short-Term Memory

## 4.2 Evaluation Methodology

As mentioned in above, this research mainly focuses on three time series analysis methods. Rather than predicting using row data, here this study uses average value of High and Low values to predict the graph. The sample graph of normalized true data is depicted in Figure 23.

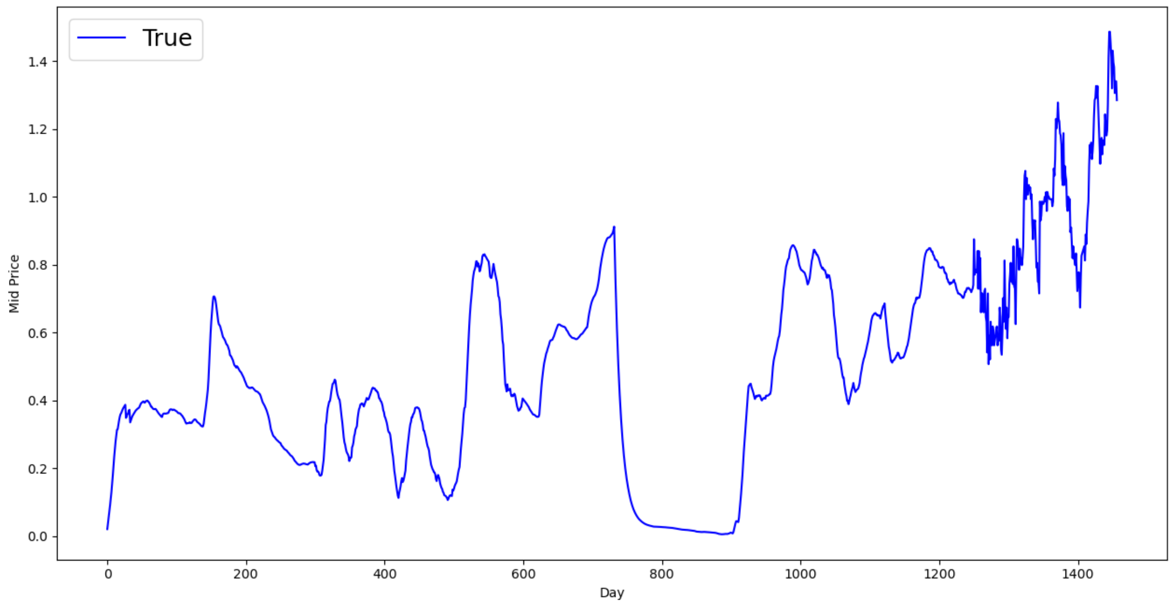


*Figure 23: Normalized Graph for the symbol AAIC*

The above graph in Figure 23 shows how the normalized price variate over time. All other companies registered under Colombo Stock Exchange shows similar behavior as the above graph. This behavior helps to test the accuracy of the created prediction model.

Data normalization is a crucial step in creating prediction model. in different time periods some of the data can have larger values and some of them might have smaller values. If there is no normalization process data which have smaller values will not make any impact on training process.

Here this study mainly focuses on two One-Step ahead prediction methods via averaging. The accuracy of the result produced by the two algorithms can be evaluated in two perspectives. Visual graphs can be used as a qualitative measure and Mean Squared Error can be used as a quantitative measure. Below Eq. 3 shows the equation of Mean Squared Error.

$$MSE = \frac{1}{N}\sum_{i=0}^{N}(f_i - y_i)^2$$

*Where N is the number of data points,* $f_i$ *is the value returned by the model and* $y_i$ *is the actual value for data point i*

*Equation 3:Equation for MSE*

### 4.2.1 Prediction via Averaging

As above mentioned this study mainly focuses on two prediction methodology via averaging. Below Figure 24 shows the output generated by using Standard (Normal) Averaging methodology.

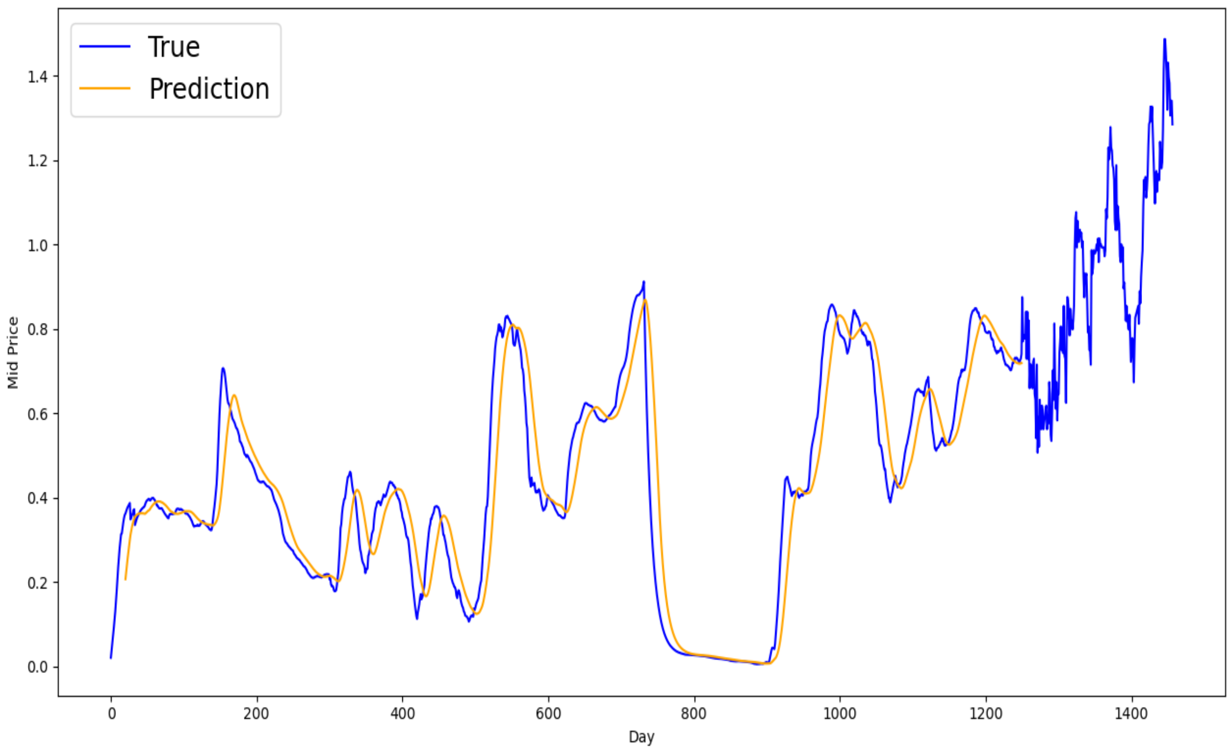


*Figure 24: Prediction made by Standard Averaging for AAIC*

Here this graph is clearly showing the prediction made by standard averaging is not completed. But both prediction and real graphs show a similar pattern. Mean Squared Error for above prediction is 0.00277. Table 1 shows the calculated Mean Squared Error for five companies listed under Colombo Stock Exchange.

| Company | AAF | ABAN | ACAP | NTB | AAIC |
|---|---|---|---|---|---|
| **MSE** | 0.00259 | 0.00299 | 0.00260 | 0.00218 | 0.00277 |

*Table 1 – MSE values calculated by standard averaging*

Above Table 1 shows that the Mean Squared Error values are close to 0.003.
The other averaging method is Exponential Moving Average method. Figure 25 illustrates the prediction made by using EMA method for the company AAIC.

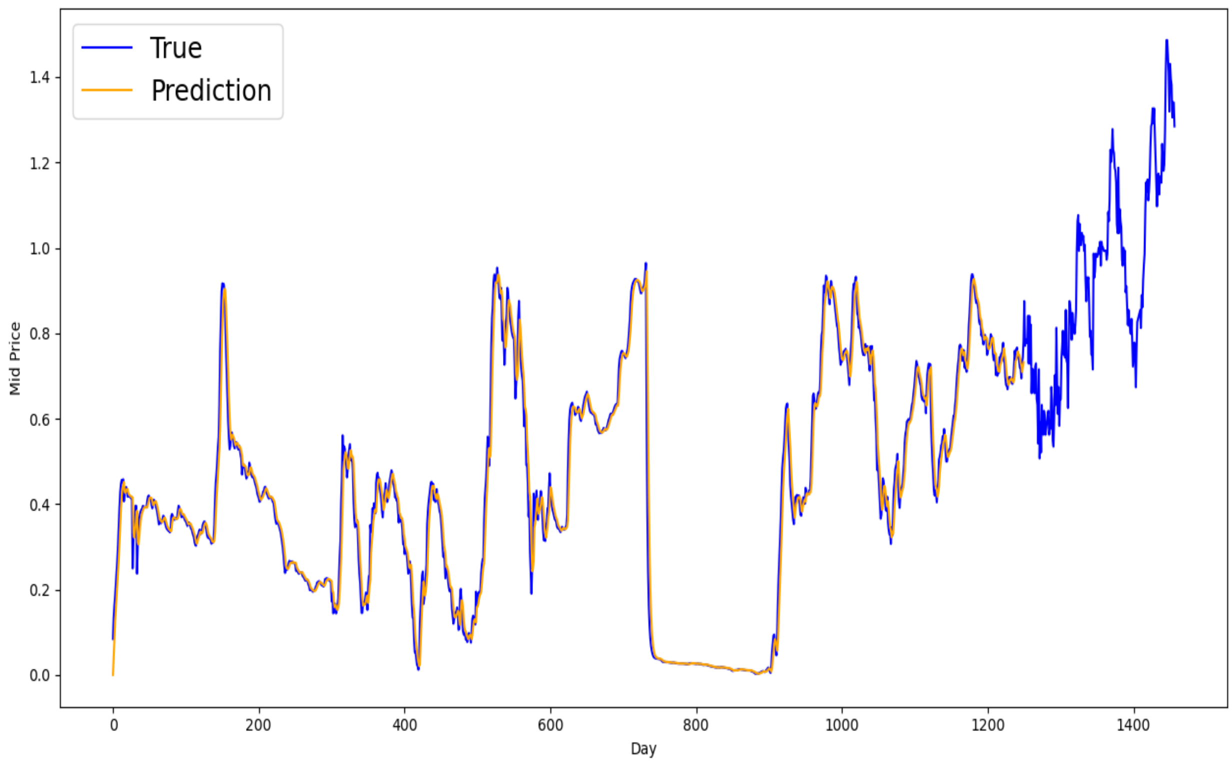


*Figure 25: Prediction made by Exponential Moving Average for AAIC*

This graph shows the prediction made by Exponential Moving Average is accurate than the prediction made by the Standard Averaging method. But EMA has not generated a completed graph as SA. Here the Mean Squared Error value for AAIC is 0.00089. It is better than Standard Averaging. Below Table 2 contains the MSE values for the five companies used in Standard Averaging.

| Company | AAF | ABAN | ACAP | NTB | AAIC |
|---|---|---|---|---|---|
| MSE | 0.00094 | 0.00047 | 0.00065 | 0.00073 | 0.00089 |

*Table 2 – MSE values from EMA*

Mean Squared Error values in the above tables and Figure 24 and Figure 25 shows that the Exponential Moving Average is better than Standard Average method.

The two graphs clearly show that an accurate prediction model can be generated by using averaging methods. Using Exponential Moving Average will help to make better predictions than Standard Averaging. But the main limitation of using Exponential Moving Average is, it relies wholly on historical data. It can be seen by the above graphs. Below is the formula for calculate EMA. The above-mentioned formula can be proved by using the below formula.

$$X_{t+1} = EMA_t = \gamma \times X_{t-1} + (1-\gamma)X_t$$

$$Where\ \gamma = \frac{N+1}{2} \text{ and N = number of days}$$

*Equation 4:EMA calculation*

Exponential Moving Average can only predict one step into the future. So, it produces same value for all the future predictions.

## 4.2.2 Long Short-Term Memory

Here this study also talks about Long Short-Term Memory. Below graph was generated by using Long Short-Term Memory method.

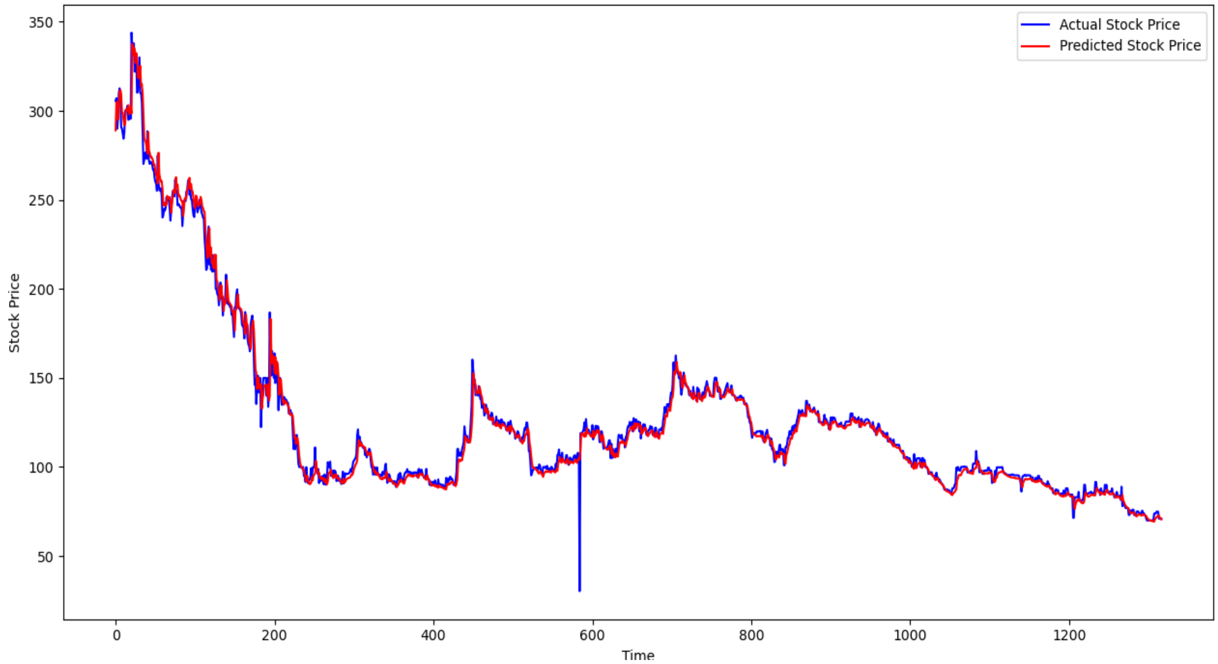


*Figure 26: Prediction made by LSTM for AAIC (without scaling data)*

Here this graph tells that LSTM can be used to make accurate and complete prediction than the Averaging methods. However, accuracy of this created model can be evaluated by looking at the graph generated by the Model.

When creating the Model mean_squarred_error has been used as the loss. This will compute the MSE. When the validation data in training, loss for train data can be seen in the console. Below Figure 27 illustrates the loss for the company AAIC.

```
Epoch 1/20
 - 8s - loss: 0.0077
Epoch 2/20
 - 7s - loss: 0.0043
Epoch 3/20
 - 7s - loss: 0.0038
Epoch 4/20
 - 7s - loss: 0.0044
Epoch 5/20
 - 7s - loss: 0.0033
Epoch 6/20
 - 7s - loss: 0.0030
Epoch 7/20
 - 7s - loss: 0.0023
Epoch 8/20
 - 7s - loss: 0.0022
Epoch 9/20
 - 7s - loss: 0.0016
Epoch 10/20
 - 7s - loss: 0.0018
Epoch 11/20
 - 7s - loss: 0.0023
Epoch 12/20
 - 7s - loss: 0.0016
Epoch 13/20
 - 7s - loss: 0.0019
Epoch 14/20
 - 7s - loss: 0.0016
Epoch 15/20
 - 7s - loss: 0.0013
Epoch 16/20
 - 7s - loss: 0.0015
Epoch 17/20
 - 7s - loss: 0.0013
Epoch 18/20
 - 7s - loss: 0.0015
Epoch 19/20
 - 7s - loss: 0.0021
Epoch 20/20
 - 7s - loss: 0.0013
```

*Figure 27: MSE for Prediction model made by LSTM for AAIC*

Chapter 5

# Conclusion

## 5.1 Conclusion

In this chapter, the author tries to describe the things that could be learned during the study. As it is described in the above chapters this study is used to find a prediction model for Stock Exchange. Before starting the research, the author has done a Literature Review about the Stock Exchange domain and the other similar prediction models. Good Literature Review helped to understand existing researchers on this problem. It also helped to improve the stock exchange domain knowledge. After doing the literature review, the author understood that this problem can be addressed in different ways.

As mentioned in the above chapters the author tried to address this research problem by developing three time series analyzing techniques. Simple averaging is the first technique used to make the prediction model. Then the author realized that simple averaging cannot produce a complete prediction. it can only be used to predict one step ahead. It relies wholly on historical data.

The Second method used in this research is Exponential Moving Average. It behaves the same as the simple averaging method. Same as the SA method EMA relies wholly on historical data. But one thing the author could realize is that the accuracy of output generated by EMA is better than the output generated by SA. EMA has good MSE values than the SA. So, the author realized that the best prediction model among SA and EMA is EMA.

The third analyzing technique used in this research is LSTM. The author realizes that the LSTM is the best technique among these three techniques. After analyzing the output generated by using the LSTM technique author realized LSTM can be used to make the prediction model as its accuracy is better than the other two techniques and it doesn’t rely wholly on historical data.

## 5.2 Future Work

The main focus of this project was to find out a prediction model to predict the stock prices in CSE. Here the author has implemented and compared three models that can be used to make predictions. But there are many other ways that can be used to make prediction models. Comparing the solution model with these prediction models will help to enhance the accuracy of the research outcome. The below list contains a few other options that can be used to make the prediction model.

- Moving Average
- Linear Regression
- k-Nearest Neighbors
- Auto ARIMA
- Prophet

In this research, quantitative attributes in CSE have been used to make the prediction model. Other than these quantitative attributes, there are qualitative attributes that can make an effect on the stock exchange. As an example, normally people sell their stocks if there are bad rumors. On another way round, people will buy stocks if there is good news. As the next step of this study, a prediction model can be implemented by using both attributes.

# References


[1] Kimberly Amadeo. "What is Stock Market?" Internet: https://www.thebalance.com/whatis-the-stock-market-how-it-works-3305893

[2] Blain Reinkensmeyer. "10 Great Ways to Learn Stock Trading in 2020" Internet: https://www.stocktrader.com/learn-stock-trading/

[3] Aditya Nawani, Himanshu Gupta, Narina Thakur, "Prediction of Market Capital for Trading Firms through Data Mining Techniques", International Journal of Computer Applications (0975 – 8887) Volume 70– No.18, May 2013

[4] Mayankkumar B Patel, Sunil R Yalamalle. "Stock Price Prediction Using Artificial Neural Network." International Journal of Innovative Research in Science, Engineering and Technology, Vol 3, pp. 13755-13762

[5] Rosalina, Hendra Jayanto. "Maximal Overlap Discrete Wavelet Transform, Graph Theory and Backpropagation Neural Network In Stock Market Forecasting." Internet: https://www.researchgate.net/publication/326862582_Maximal_Overlap_Discrete_Wavelet_Transform_Graph_Theory_And_Backpropagation_Neural_Network_In_Stock_Market_Forecasting.

[6] Li Zhang, Fulin Wung, Bing Xu, Wenyu Chi, Qiongya Wang, Ting Sun. Prediction of Stock Prices Based on LM-BP Neural Network and the Estimation of Overfitting point by RDCI." in Proceedings of the Natural Computing Application Forum, 2017.

[7] Applications of ANNs in Stock Market Prediction: A Survey. neha Soni / International Journal of Computer Science & Engineering Technology (IJCSET)

[8] A.J.P Samarawickrama, T.G.I Fernando. "A recurrent neural network approach in redicting daily stock prices an application to the Sri Lankan stock market." In Proceeding of the IEEE Conference, 2017.

[9] Arosha Senanayake, J Jananaththanam. "Efficient Neural Stock Market Prediction System.", Internet: http://www.icter.org/conference/icter2016/sites/default/files/icter/IITC-2002p20.pdf.

[10] Jingi Tang, Xiong Chen. "Stock Market Prediction Based on Historic Prices and News Titles." in Proceedings of International Conference on Machine Learning Technologies, 2018.

[11] Yue-Gang Song, Yu-Long Zhou, Ren-Jie Han. “Neural Networks for Stock Price Prediction.”, 2018, Internet: https://arxiv.org/abs/1805.11317.

[12] Jaromir Vrbka, Zuzana Rowland. “Stock price development forecasting using neural networks.” Innovative Economic Symposium, Vol 39.

[13] Adam Milton, “Simple, Exponential, and Weighted Moving Averages”, Internet: https://www.thebalance.com/simple-exponential-and-weighted-moving-averages-1031196

[14] Edwards, R.; Magee, J, “Technical Analysis of Stock Trends. New York”, New York Institute of Finance, 1992.

[15] Girdzijauskas, S., et al., “Ekonominių burbulų susidarymas ir galimybės jų išvengti [Formation of economic bubbles: causes and possible preventions]”, 2009

[16] Ehlers, J., & Way, “Zero lag. Mesasoftwareweb, technical papers”, 2010, Internet: http://www. mesasoftware.com/papers/ZeroLag.pdf

[17] James, Jr., F.E.: “Monthly Moving Averages—An Effective Investment Tool?" Journal of Financial and Quantitative Analysis, September 1968, pp. 315–326.

[18] All Answers Ltd, “Technical Analysis of Markets.”, November 2018, Internet: https://ukdiss.com/litreview/the-trend-of-technical-analysis.php

[19] Murphy, J. J. “Technical analysis of the financial markets”, 1st edition, United States of America, New York Institute of Finance, Internet:- http://jultika.oulu.fi/files/nbnfioulu-201605121716.pdf

[20] B Uma Devi, D Sundar, P Ali. “An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50.” International Journal of Data Mining & Knowledge Management Process, Vol 3, 2017, pp.65-78.

[21] Mahantesh C Angadi, Amogh P Kulkarni. “Time Series Data Analysis for Stock Market Prediction using Data Mining Techniques with R.” International Journal of Advanced Research in Computer Science, Vol 6, 2015, pp.104-108.

[22] Ayodele A. Abebiyi, Aderami O. Adewumi, Charles K Ayo. “Stock Pice Prediction sing the ARIMA Model.” in Proceedings of the UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 2014.

[23] Kamalakannan J, Indrani Senguptha, Snehaa Chaudhury. “Stock Market Prediction Using Time Series Analysis.” in Proceedings of the IADS International Conference on Computing, Communication & Data Engineering, 2018.

[24] Mohamed Ashik, Senthamarai Kannan K. “Forecasting National Stock Price Using ARIMA Model.” Global and Stochastic Analysis, Vol 4, 2017 pp.77-81.

[25] Tina Ding, Venessa Fang, Daniel Zuo. “Stock Market Prediction based on Time Series Data and Market Sentiment.” Internet: http://murphy.wot.eecs.northwestern.edu/~pzu918/EECS349/final_dZuo_tDing_vFang.pdf.

[26] A. V Jadhav, K.B Kamble. “Prediction of Stock Price in Oil Sectors using ARIMA Model.” International Journal of Mathematics Trends and Technology, Vol 51, 2017 pp. 266-270.

[27] Deepthi Praveenlal Kuttichira. “Stock Price Prediction Using ARIMA.” International Journal of Engineering Technology Science and Research, Vol 4, 2017 pp. 676-678.

[28] Adi Bronshtein. “Train/Test Split and Cross Validation in Python” Internet: https://towardsdatascience.com/train-test-split-and-cross-validation-in-python80b61beca4b6

[29] Jeff Heaton. “Introduction to Neural Networks in Java”, 2005