CHAPTER 1

INTRODUCTION

1.1 Background

Stock Market prediction and analysis is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. Stock market is the important part of economy of the country and plays a vital role in the growth of the industry and commerce of the country that eventually affects the economy of the country. Both investors and industry are involved in stock market and wants to know whether some stock will rise or fall over certain period of time. The stock market is the primary source for any company to raise funds for business expansions. It is based on the concept of demand and supply. If the demand for a company's stock is higher, then the company share price increases and if the demand for company's stock is low then the company share price decrease.

Another motivation for research in this field is that it possesses many theoretical and experimental challenges. The most important of these is the Efficient Market Hypothesis(EMH), the hypothesis says that in an efficient market, stock market prices fully reflect available information about the market and its constituents and thus any opportunity of earning excess profit ceases to exist. One of the example of big exchange is New York Stock Exchange.

The performance of stock markets is measured on a daily basis by some key indicators such as 'share index', which is a measure of the performance of some stocks picked from the different sectors of the market. Such an index is important in not only gauging the performance of trades in the stock exchange but also the economic performance of the particular country as a whole.

Shareholders however do not directly execute the trade, nor is there any meeting between buyers and sellers for negotiations. Shareholders trade by giving instructions to their Stockbrokers, who in turn execute the orders. Stockbrokers usually also advise clients on where to trade. In their advisory role, some Stockbrokers base their advice on the fundamentals of the various stocks or undertake technical analysis. However, none of these predictive methods have assurance of profit as they usually just indicate a future trend and a likely up or down price movement and not the real expected future stock price. Stockbrokers need to be empowered, through better predictive tools, to enable them have some capability to provide

the best advice to their clients. A predictive tool that Stockbrokers can use to guide on exact price movements, as a basis of investment, is therefore desirable. This can be an artificial intelligence (AI) system based on neural networks.

1.2 Motivation

This project is done as a partial fulfillment of requirements for the award of the degree of Bachelor of Engineering in Computer Engineering (Seventh semester). Besides this we wanted to do a project which implements an Artificial Intelligence (AI), data science, machine learning and finance. Stock market analysis and prediction could be done implementing all of those so this fact motivated us on selecting this project.

1.3 Statement of Problems

Trading in shares is big business in many economies. Due to the changing demands and supply stock prices regularly fluctuate. Many people believe that the stock market prediction is an impossible task as it is not easy to follow patterns and apprehensive for investing in stock market. The investors generally refer advice from the stockbrokers who mostly uses technical, fundamental or time series analysis in trying to predict the stock market. These prevalent methods show a trend on future movement and not the likely trade price for any stock in future.

Stockbrokers do not seem to have any intelligent tool that can help them advise clients on which stocks are suitable for any buy or sale trade. It is therefore desirable to have a tool that does not just point a direction of price movement, but also indicates the most likely price value of the stock itself. An Artificial Neural Network (ANN) model that is well tuned with the appropriate parameters can be used to develop such a predictive tool.

1.4 Objective

To analyse the effectiveness of the implemented prediction algorithms on stock market data and provide general insight on this data to user through visualization.

1.5 Scope and Limitation

Scope:

This system can be used as a platform by the user to know about the stock value and its prediction. It’s a web application that can used for various tasks and applications such as decision making, internet marketing and information gathering about the stock market value.

Limitation:

This system has its own limitations. The various limitations of this project due to various circumstances are as follows.

- Since, there are many indeterminate parameters that directly affect stock market, each and every one of them cannot be taken into account. So, our model only depends on the relationship of our selected parameters with the share price.
- At present, this system only performs analysis and prediction of stock market value that are available in the Yahoo Finance so it is not applicable for Nepal Stock market.
- This system is limited to only certain users that have knowledge of stock market.
- Our prediction model can predict accurate to some stock but not for all.

1.6 Application

- Stock brokers company.
- Applicable for technical analysis as well as technology analysis of stock markets.

CHAPTER 2
LITERATURE REVIEW

Generally speaking, stock prediction methodologies are consisting of three main categories as fundamental analysis, technical analysis (charting) and technological methods. On the one hand, fundamental analysis believes that the stock trends are highly related to the company that underlies. They evaluate a company's credibility of its accounts as well as its past performance using performance ratios, such as P/E ratio, to aid the analysis process. Technical analysis, on the other hand, predicts the future price of a stock based solely on the trends of the past prices. During the analysis, numerous patterns such as the "head and shoulder" and "cup and saucer", as well as statistical techniques such as the exponential moving average (EMA) are adopted by the technical analysts. Last but not least, with the advent of the digital computer, stock market prediction has moved in to the technological realm.

In the technological methods category, the most promising technique involves the use of artificial neural networks (ANNs). For stock prediction, the mostly used type of ANNs is the feed forward network with backward propagation algorithm in the weight training phase. In addition, another form of ANN that is especially suitable for stock prediction is the recurrent neural network (RNN).

Recurrent Neural Network (RNN)

A Recurrent neural network (RNN) is a type of advanced artificial neural network (ANN) that involves directed cycles in memory. One aspect of recurrent neural networks is the ability to build on earlier types of networks with fixed-size input vectors and output vectors.

The use of recurrent neural networks is often related to deep learning and the use of sequences to evolve models that simulate the neural activity in the human brain. In terms of practical application, RNNs have been an active area of focus for many professionals for uses like image processing, language processing, and even models that add characters to text one at a time, music generation and stock market prediction and analysis.

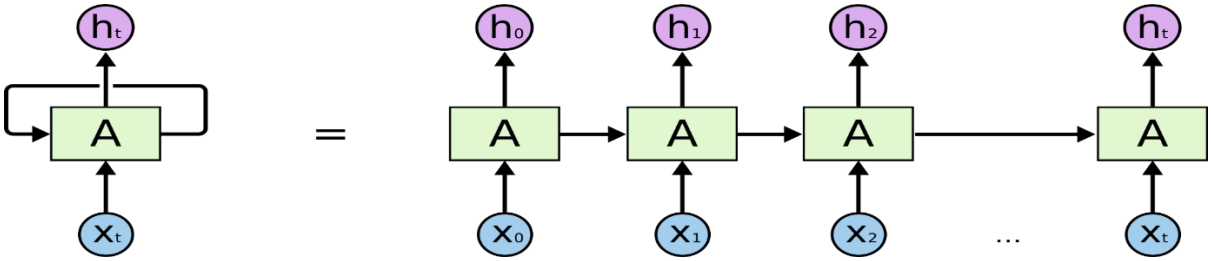

Figure 2.1.1: Recurrent Neural Network

2.1 Related Work

The current state-of-the-art result in stock prediction using RNN is from Armando B.'s Financial Market Time Series Prediction with Recurrent Neural in 2012. [4] They used an Echo State Networks (ESNs) which is also a branch of RNN to try to capture the inherent nature of stock data with feature vector included the current and 5-day history of the stock price, the 5, 10, 15, and 20 day moving averages, volume. As for data source, the S&P500 index is use. And the test error, measured by Root Mean Square Error (RMSE), for the ESN is 0.0027.

Wilson and Sharda [2] studied prediction firm bankruptcy using neural networks and classical multiple discriminant analysis, where neural networks performed significantly better than multiple discriminant analysis.

Min and Lee were doing prediction of bankruptcy using machine learning. They evaluated methods based on Support Vector Machine, multiple discriminant analysis, logistic regression analysis, and three-layer fully connected back-propagation neural networks. Their results indicated that support vector machines outperformed other approaches.

In the technological methods category, there are other popular trends. Two most significant ones are the use of Text Mining together with Machine Learning algorithms and the use of new statistical analysis tools of complexity theory (researchers at the New England Complex Systems Institute (NECSI) performed research on predicting stock market crashes.)

Tsai and Wang [3] did a research where they tried to predict stock prices by using ensemble learning, composed of decision trees and artificial neural networks. They created dataset from Taiwanese stock market data, taking into account fundamental indexes, technical indexes, and macroeconomic indexes. The performance of Decision Tree + Artificial Neural Network trained on Taiwan stock exchange data showed F score performance of 77%. Single algorithms showed F-score performance up to 67%.

There are many tools and software available out there that provide forecasting of stock market entities, share quantity and share value for a given financial organization. Most of them claim to predict the stock market with near to 100% accuracy but the opinions from the users vary. Some of the popular tools and software with their methodologies are mentioned as follows.

2.2. inteliCharts Predictive Stock Market Analytics

It is a quantitative modelling tool used for financial time series forecasting. The system is adaptive in its core as it learns the patterns and geometrical relationships defined by historical time series data points, which are unique for each individual stock, index, or another financial instrument

2.3. Markettrak

Its stock market forecast system consists of two major parts: an extensive database and a forecast model. The forecast model reads the database and then makes a prediction of where the market is headed. From this prediction, it determines a trading position for the Dow Diamonds or the SP500 Spiders. The database and forecast are updated daily at the close of trading. It uses a neural network model in combination with a genetic algorithm to calculate the SP500 forecast.

2.4. Stock-Forecasting.com

An innovative stoke price prediction web-based software for stock market forecasting and analysis. The artificial intelligence www.stock-forecasting.com software is based on neural network technology, advanced statistical methods and non-periodic stock price wave analysis. The Stock-Forecasting software predicts stock prices, generates trading "Buy-Hold-Sell" signals, computes the most profitable company to invest in and analyses the accuracy of predictions.

2.5. Nepal Stock Information

Nepal stock information is an interactive web application which contains all the data’s and graph of the major companies of Nepal stocks. This web application focuses mainly on the technical analysis of stock market rather than technological. It contains many important features and parameters for making a technical analysis of stock markets of Nepal.

CHAPTER 3

PROJECT MANAGEMENT

3.1 Group Member Details

For this project, we have a group of four members:

1. Ajay Shrestha (710301)

2. Anish Karmi (710303)

3. Sajan Basnet (710336)

4. Sudan Krishna Shrestha (710345)

5. Sushant Ghimire (710347)

3.2 Gantt Chart

The work breakdown structure for the proposed project is as given below.

Weeks / Works	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Project identification	■									
Analysis		■								
Design			■	■						
Coding and testing					■	■	■			
Report making								■	■	
Documentation		■	■	■	■	■	■	■	■	■

Table 3.2.1: Gantt chart

3.3 Feasibility Study

Simply put, stock market cannot be accurately predicted. The future, like any complex problem, has far too many variables to be predicted. The stock market is a place where buyers and sellers converge. When there are more buyers than sellers, the price increases. When there are more sellers than buyers, the price decreases. So, there is a factor which causes people to buy and sell. It has more to do with emotion than logic.

Because emotion is unpredictable, stock market movements will be unpredictable. It's futile to try to predict where markets are going. They are designed to be unpredictable.

There are some fundamental financial indicators by which a company's stock value can be estimated. Some of the indicators and factors are: Price-to-Earning (P/E) Ratio, Price-to-Earning Growth (PEG) Ratio, Price-to-Sales (P/S) Ratio, Price/Cash Flow (P/CF) Ratio, Price-to-Book Value (P/BV) Ratio and Debt-to-Equity Ratio, OHLC charts. Some of the parameters are available and accessible on the web but all of them aren't. So we are confined to use the variables that are available to us.

The proposed system will not always produce accurate results since it does not account for the human behaviours. Factors like change in company's leadership, internal matters, strikes, protests, natural disasters, change in the authority cannot be taken into account for relating it to the change in Stock market by the machine.

The objective of the system is to give an approximate idea of where the stock market might be headed. It does not give a long term forecasting of a stock value. There are way too many reasons to acknowledge for the long term output of a current stock. Many things and parameters may affect it on the way due to which long term forecasting is just not feasible.

3.4 Requirement Analysis

After the extensive analysis of the problems in the system, we are familiarized with the requirement that the current system needs. The requirement that the system needs is categorized into the functional and non-functional requirements. These requirements are listed below:

3.4.1 Functional Requirements

Functional requirement are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this, the functional requirements that the system must require are as follows:

- The system should be able to generate an approximate share price.

- The system should collect accurate data from the Yahoo Finance website in consistent manner.
- The system should give its user with an ability to choose the desired stock market name and symbols.
- The system should be able visualize the chosen stock market data in the form of different graphs.

3.4.2 Non-Functional Requirements

Non-functional requirement is a description of features, characteristics and attribute of the system as well as any constraints that may limit the boundaries of the proposed system. The non-functional requirements are essentially based on the performance, information, economy, control and security efficiency and services. Based on these the non-functional requirements are as follows:

- The system should provide better accuracy.
- The system should have user friendly interface.
- The system should be reliable with an efficient performance.

CHAPTER 4

METHODOLOGY

4.1 Block Diagram

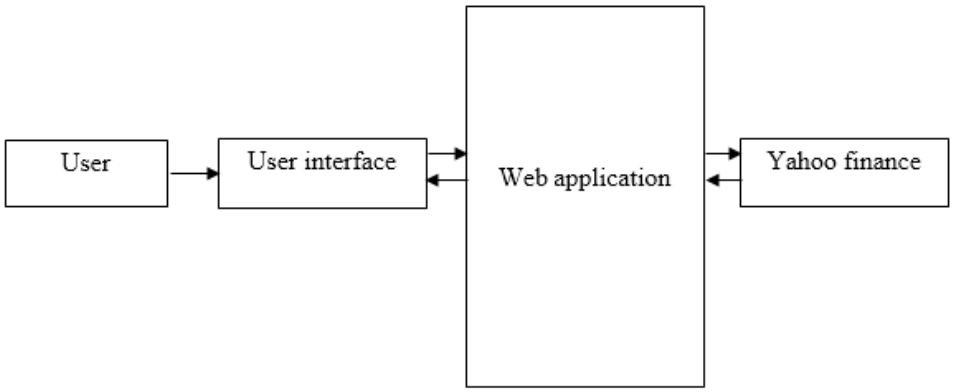

Figure 4.1.1 Block diagram

As seen in above block diagram, users can view various stocks. They will be able to check its trading history. The trading data is pulled from Yahoo finance. The data is then used to predict the opening price of the stock.

4.2 Program flow

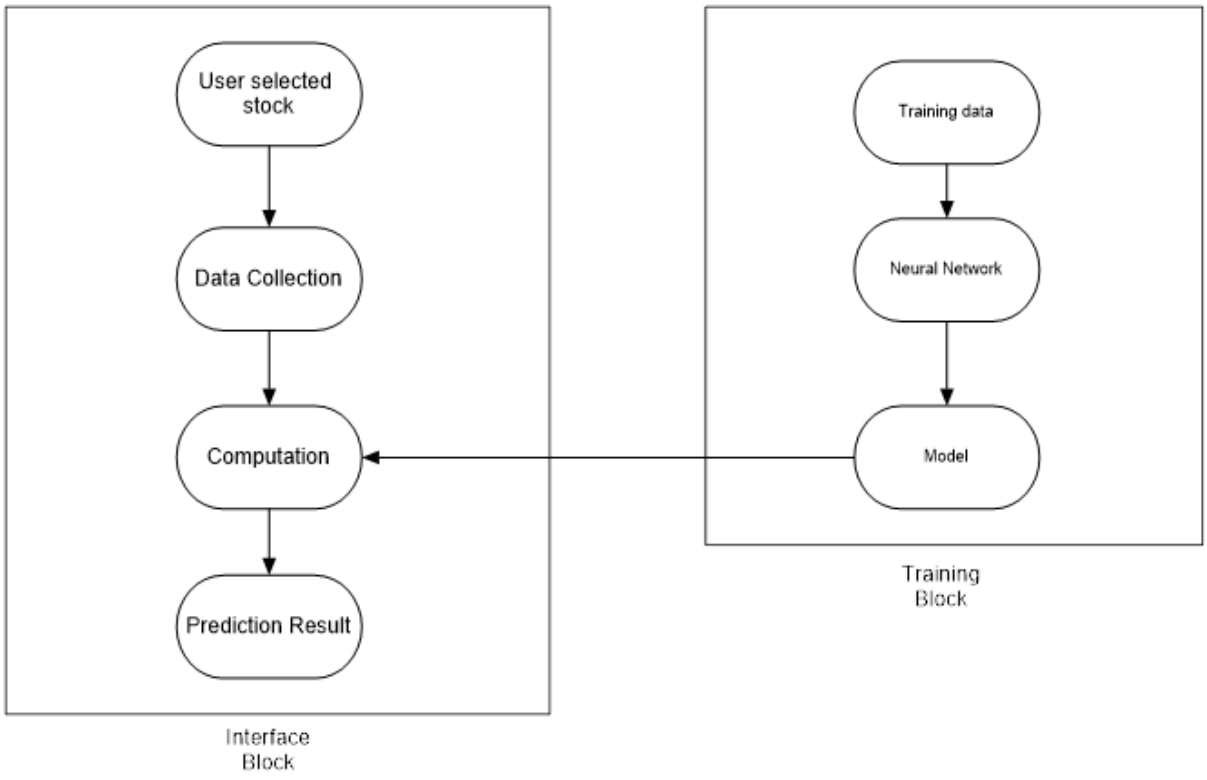

Figure 4.2.1 Program flow

4.3 Use Case Diagram

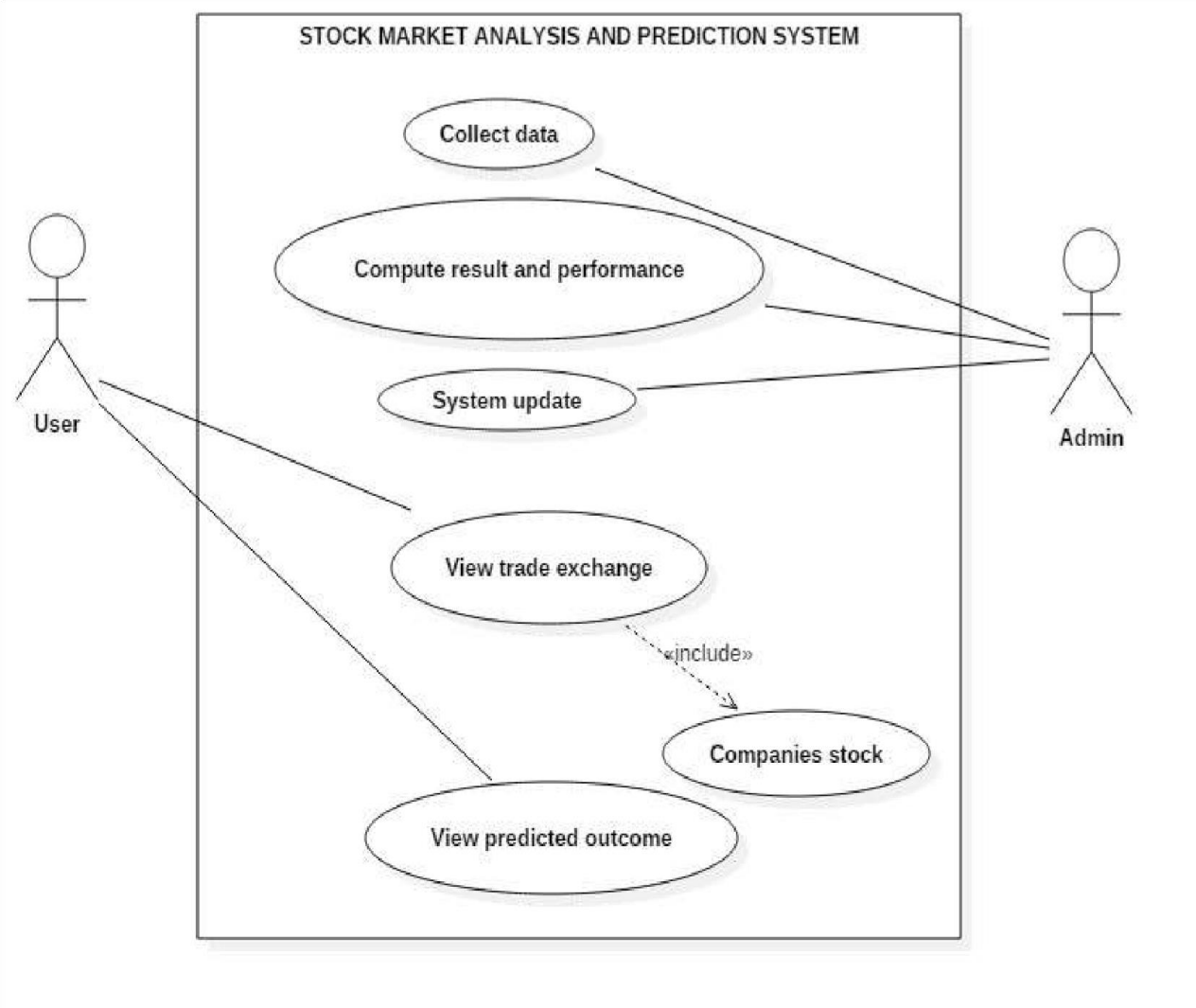

Figure 4.3.1 Use case diagram

4.4 Data source

Historical data is very important for the prediction of the stock prices. It helps to gain a more accurate prediction. The trading data is extracted from the Yahoo finance site (https://finance.yahoo.com). It provides stock trading data with all the required details such as: Open, High, Low, Close, Volume and Adjusted close. The data is extracted on a daily basis.

4.5 Recurrent Neural Network

A recurrent neural network (RNN) is a class of artificial neural network where connections between units form a directed cycle. This allows it to exhibit dynamic temporal behaviour. Unlike feed forward neural networks, RNNs can use their internal memory to process arbitrary sequences of inputs.

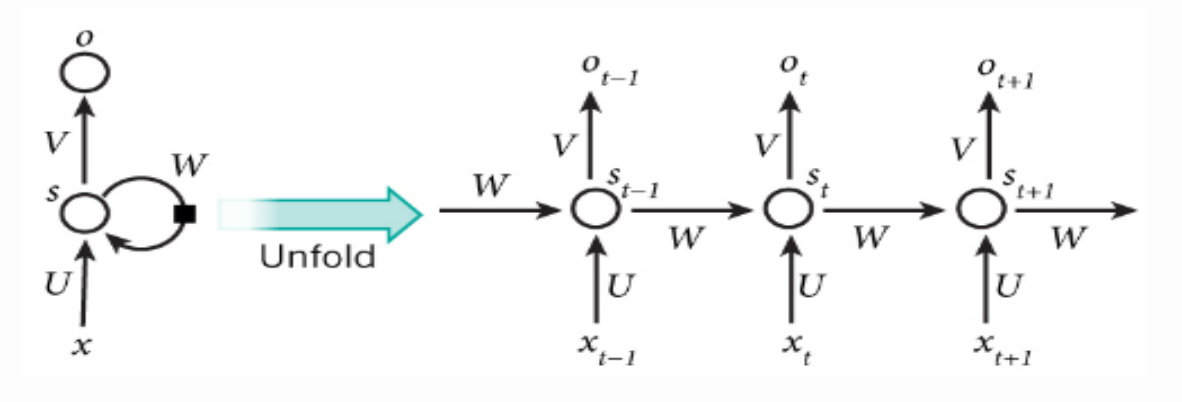

Figure 4.5.1: Recurrent neural network

RNNs are very apt for sequence classification problems and the reason they're so good at this is that they're able to retain important data from the previous inputs and use that information to modify the current output. If the sequences are quite long, the gradients (values calculated to tune the network) computed during their training (back propagation) either vanish (multiplication of many 0 < values < 1) or explode (multiplication of many large values) causing it to train very slowly.

4.6 Long Short-Term Memory

Long Short Term Memory networks "LSTMs" are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hoch Reiter & Schmidhuber (1997), and were refined and popularized by many people in following work. They work tremendously well on a large variety of problems, and are now widely used.

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn. All recurrent neural networks have the form of a chain of repeating modules of neural network.

Follow algorithm is used to build a LSTM cell network.

1. LSTM cells are defined using equations. Each cell consists of 3 gates namely input, forget and output.
2. After definition of gates in cell, forget gate is used to throw away some information. 1 represents "completely keep this" and 0 represents "completely get rid of this".
3. Then new information to store is decided and input layer is used to update the values.
4. Finally, we need to decide what we're going to output. This output will be based on our cell state but will be a filtered version.

4.7. Data Normalization

The data is normalized before being input to the ANN. The input vectors of the training data are normalized such that all the features are zero-mean and unit variance. The target values are normalized using min-max function such that all the values are converted into the values within the range of 0 to 1. The minimum value is represented by 0 and the maximum value is represented by 1.

$$z = \frac{x - min(x)}{max(x) - min(x)}$$

4.8. Activation Function

We use log sigmoid function as the activation function at both hidden layer and output layer. A sigmoid function is a mathematical function having a characteristic "S" shaped curve or sigmoid curve which is given by:

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}.$$

It transforms linear inputs to nonlinear outputs. It bounds output to between 0 and 1 so that it can be interpreted as a probability and it also makes computation easier than arbitrary activation functions. The reason we choose this function as activation function is because it gives logistic neurons real-valued output that is a smooth and bounded function of their total input. It also has the added benefit of having nice derivatives which make learning the weights of a neural network easier.

4.9. Training Parameters

In order to select optimal parameters for the neural network, simulation is carried out. A model of a neural network is constructed and simulated using simulation tool. Test runs are carried out and the model yielding the best accuracy is selected for implementation. The best model so far has the following parameters:

Learning rate: 0.10

Total Layers: 5

Input Neurons: 1

Hidden Layer Neurons: 3

Output Neurons: 1

Activation Function: Log Sigmoid

Number of epochs used: 1000

4.10. Data format

Here is a brief description about the data format found in stocks.

4.10.1. Highest Value

It is the highest value the share price of a company that has reached in the previous day.

4.10.2. Lowest Value

Similar to highest value, it is the highest value the share price of a company that has reached in the previous day.

4.10.3. Share Volume

Share volume can be calculated in two different types the daily share volume and the monthly share volume. The total number of share is sold in a particular day is called daily share volume. In monthly share volume is the sum of the trading volumes during that month.

4.10.4. Opening price

Opening price refers to the first price the stock starts in the trading day.

4.10.5. Closing price

Closing price generally refers to the last price at which a stock trades during a regular trading session. For NEPSE, regular trading sessions run from Sunday - Thursday (11:00 AM. to 3:00 PM) and Friday(12:00 pm to 1:00 pm).

CHAPTER 5

RESULT AND DISCUSSION

5.1 Neural Network Training:

The neural network is a RNN in which we use LSTM Algorithm to train our network. Our current model, we used 3 hidden layers with 400 units in each layer. We used The Standard & Poor's 500, often abbreviated as the S&P 500, stock data as our training data set. This data set contains data from 3/1/1950 to 11/5/2018 which sums up to 17201 data. We train our network with many epochs or iterations. The result of the training is shown below:

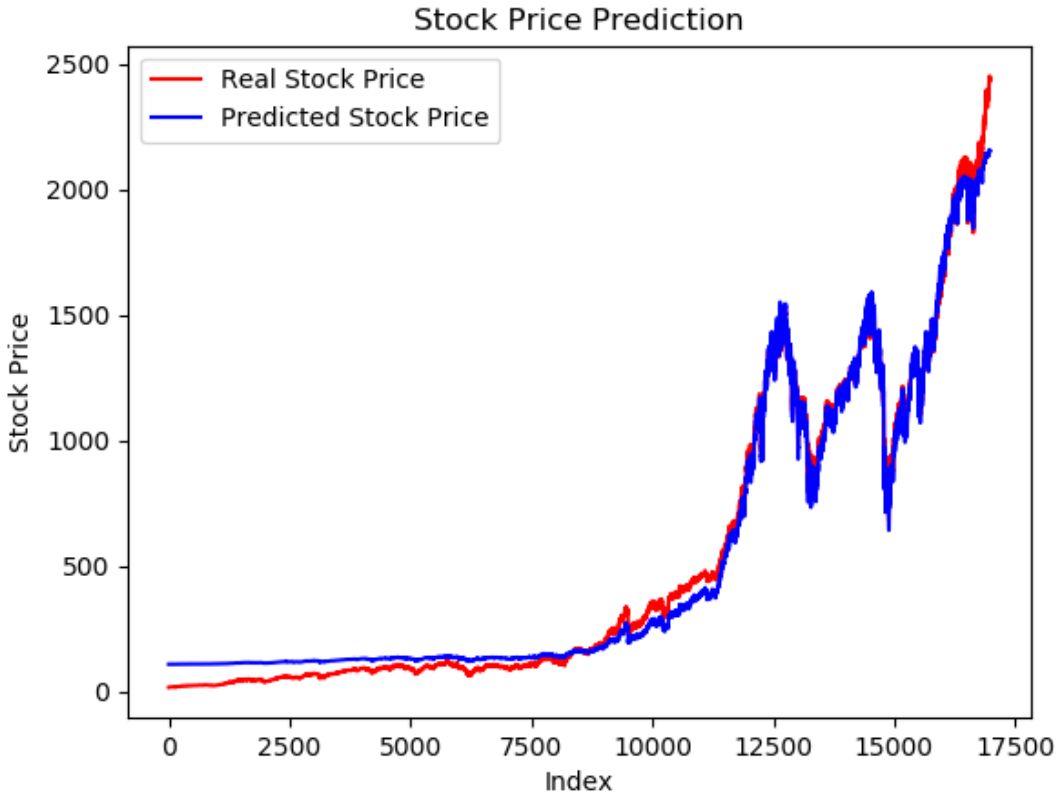

Figure 5.2.1: Training with 100 epochs

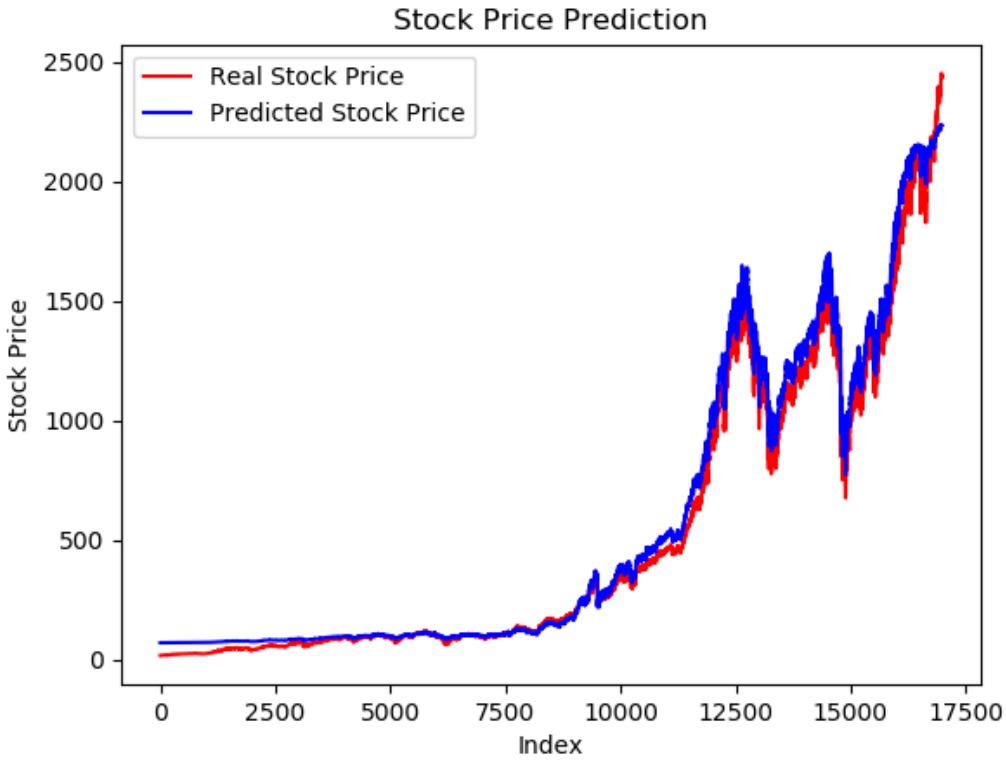

Figure 5.2.2: Training with 500 epochs

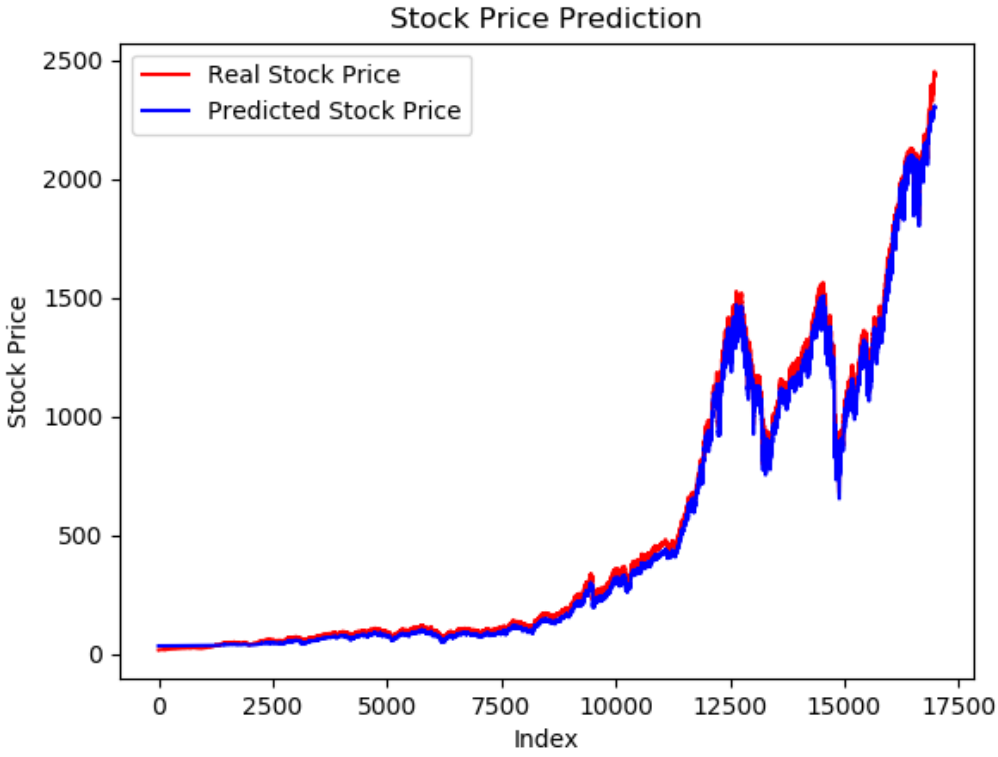

Figure 5.2.3: Training with 1000 epochs

5.2 Testing

5.2.1 Stages of Testing Process

There are five stages of system where system components are tested, the integrated system is tested with stock data. The five levels of system testing include unit testing, module testing, subsystem testing, system testing and acceptance testing.

1. Unit Testing

Unit testing is a level of software testing where individual units/ components of a software are tested. The purpose is to validate that each unit of the software performs as designed. A unit is the smallest testable part of any software. It usually has one or a few inputs and usually a single output. It is also called as component testing.

2. Module Testing

The related collection of dependent components are tested in module testing without other system module.

3. Sub-system Testing

Modules are integrated into sub-system are tested. The main objective of this testing is to interface testing to detect interface error or mismatches.

4. System Testing

System testing is the process of testing an integrated system to verify that it meets specified requirements. It also includes the validation of functional and non-functional requirements.

5. Acceptance Testing

Acceptance testing is a formal testing with respect to user needs, requirements, and business processes conducted to determine whether or not a system satisfies the acceptance criteria and to enable the user, users or other authorized entity to determine whether or not to accept the system. This testing makes sure that the system works correctly for intended user in normal working environment. There are two types of acceptance testing. They are:

i. Alpha Testing

This type of software testing involves the testing of system by the user under the supervision of developer at the developer's site.

ii. Beta Testing

This is the type of acceptance testing where the complete software id tested by the user at own site without the presence of developer.

5.3 Test cases:

Test case 1:

Company Name	Range of data used	Number of data	Accuracy (5% significance)	Correlation coefficient
Apple Inc. (AAPL)	17-08-2001 to 14-07-2017	4000	89.73	0.99775
Bank of America Corporation (BAC)	3-17-1980 to 14-07-2017	9607	89.29	0.99839
Tesla, Inc. (TSLA)	6-29-2010 to 20-04-2018	1966	93.15	0.99854
Facebook, Inc. (FB)	5-18-2012 to 20-04-2018	1489	94.91	0.99815
Amazon.com, Inc. (AMZN)	5-15-1997 to 20-04-2018	5267	90.04	0.99844

Table 5.3.1: Overall accuracy tests

Test case 2:

Stock	Actual		Predicted	
	1st day	2nd day	1st day	2nd day
Apple Inc.	187.63	187.16	186.45	186.91
Bank of America Corporation	30.81	30.26	30.56	30.30
Tesla, Inc.	284.57	276.82	281.64	274.88
Facebook, Inc.	183.76	182.68	181.47	181.98
Amazon.com, Inc.	1581.76	1574.37	1580.54	1575.91

Table 5.3.2: Actual vs Predicted value

CHAPTER 6

CONCLUSION AND RECOMMENDATION

6.1 Conclusion

We implement the application of Recurrent Neural Network with LSTM to the task of stock market prediction. Our initial analysis show significant correlation between different input parameter. The result obtained in test cases was fairly accurate. As is indent from table 5.3.1, 5.3.2, the prediction is fairly accurate unless there is huge and sudden variation in the actual data. On other hand, this also proves the hypotheses that stock market are actually unpredictable.

After the phase of data collection prediction, the result will be displayed to users in the form web pages. The web pages shows graphs like line chart, bar chart and candle stick graph.

6.2 Future Recommendation

- Addition of different stock indicators which helps to show the market conditions like rising, up trend, down trend, etc.
- This system only shows the predicted stock values. For future work, we can add interface to buy or sell the stock directly.
- Addition of login system which can show all the stock exchange performed by the user and allows user to give feedback.

REFERENCES

[1]Investopedia, https://www.investopedia.com/terms/s/stockmarket.asp [10 March 2018]

[2]R. Wilson and R. Sharda, "Bankruptcy prediction using neural networks", *Decision Support Systems*, vol. 11, no. 5, pp. 545-557, 1994.

[3]K. Tsai and J. Wang, "External technology sourcing and innovation performance in LMT sectors", *Research Policy*, vol. 38, no. 3, pp. 518-526, 2009.

[4]Armando B., Sam F., Rohit P. Financial Market Time Series Prediction with Recurrent Neural Networks [20 April 2018]

Yahoo finance, https://finance.yahoo.com [25 April 2018]

Nepal stock information, http://nepalstockinfo.com/ [27 April 2018]

APPENDIX

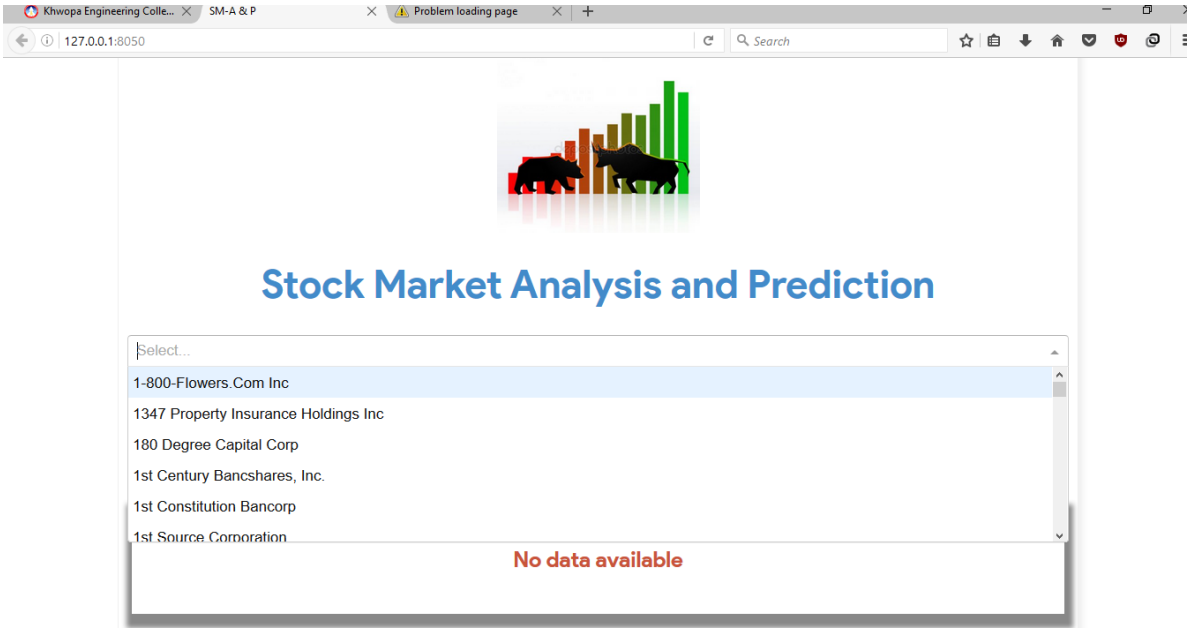

Figure 1: Homepage

Figure 2: Stock closing price with prediction

Figure 3: Stock closing price without prediction

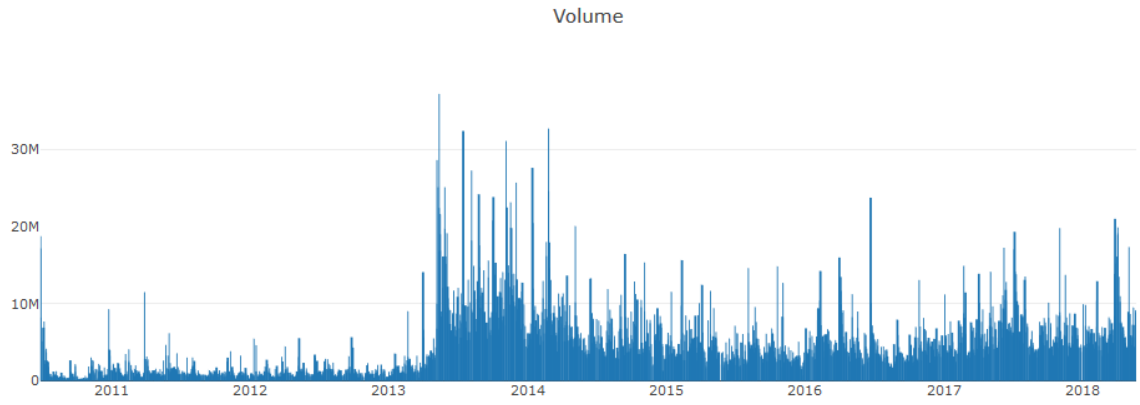

Figure 4: Stock volume

Figure 5: 30 days closing price with prediction

Figure 6: Future prediction

Figure 7: Candle stick graph