GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

STOCK PRICE PREDICTION USING DEEP LEARNING

A Project Report of Capstone Project - 2

Submitted by

Ujjwal Kumar

(1613101789 / 16SCSE101145)

in partial fulfilment for the award of the degree

of

Bachelor of Technology

IN

Computer Science and Engineering

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

Under the supervision of

Mr SAMSON EBENEZAR

Assistant Professor

APRIL / MAY -2020

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report **"STOCK PRICE PREDICTION BY USING DEEP LEARNING"** is the bonafide work of **"UJJWAL KUMAR(1613101789)"** who carried out the project work under my supervision.

SIGNATURE OF HEAD	**SIGNATURE OF SUPERVISOR**
Dr. MUNISH SHABARWAL,	**Mr Samson Ebenezar,**
PhD (Management), PhD (CS)	**Assistant Professor**
Professor & Dean,	**School of Computer Science &**
School of Computer Science &	**Engineering**

ACKNOWLEDGEMENT:

This is an excellent opportunity to acknowledge and to thanks, all those persons without whose support and help this project would be impossible. We might prefer to add some heartfelt words for those who were a part of this project in numerous ways.

I would prefer to because of my project guide **Mr Samson Ebenezar**, for her indefatigable guidance, valuable suggestion, moral support, constant encouragement, and contribution of your time for the successful completion of project work. I'm very grateful to him, for providing all the facilities needed during the project development. At the outset, I sincerely thank all faculty members of my institution **GALGOTIAS UNIVERSITY** for their extra effort to create our session online and inspire all ideas.

I would prefer to thank all those that helped me directly or indirectly. Last but not the least, I'd prefer to acknowledge the continuing support of my friends, whose patience and encouragement during these long days and nights are paramount in making this project a reality.

THANK YOU.

DECLARATION:

I hereby declare that this submission is my very own work which, to the simplest of my knowledge and belief, it contains no material previously published or written by another person nor material which to a considerable extent has been accepted for the award of the other degree or diploma of the university or other institute of upper learning, except where due acknowledgment has been made within the text.

I inform that every data used in this report if it's taken from any site is clearly referenced under the reference section.

SIGNATURE

Ujjwal kumar

16SCSE101145

Date: 17-may-2020

TABLE OF CONTENTS

ABSTRACT

In a financially volatile market, as the stock market, it is important to have a very precise prediction of a future trend. Because of the financial crisis and scoring profits, it is mandatory to have a secure prediction of the values of the stocks. Predicting a non-linear signal requires advanced algorithms of machine learning. The literature contains studies with different machine learning algorithms such as ANN (artificial neural networks) with different feature selection. The results of this study will show that the algorithm of classification SVM (Support Vector Machines) with the help of feature selection PCA (Principal component analysis) will have the success of making a profit.

This paper systematically reviews studies that forecast stock market prediction and analysis using different methods. We extract and synthesise 150 research papers, we found different methods by which we can predict the stock market condition and by which investors can easily invest their money in the stock market. Some of those methods are:-

(1) Methods using FUZZY LOGIC methods and (2) Fundamental Analysis

(3) Artificial Neural Network (4) Data Mining

(5) Hidden Markov Model .

FUZZY TIME SERIES forecasting methods, which have been widely studied in recent years, do not require constraints as found in conventional approaches. On the other hand, most of the time series encountered in real life should be considered as fuzzy time series due to the vagueness that they contain. Although numerous methods have been proposed for the analysis of time series in the literature, these methods fail to forecast seasonal fuzzy time series. Neuro-Fuzzy system used to predict the stock market fluctuation. NEURAL NETWORK and NEURO-FUZZY systems are identified to be the leading machine learning techniques in the stock market index prediction area. The Traditional techniques do not cover all the possible relation of the stock price fluctuations. There are new approaches known in-depth of an analysis of stock price variations.

NN and MARKOV Model can be used exclusively in the finance markets and forecasting of stock price.Forecasting stock return is an important financial subject that has attracted researchers' attention for many years. It involves an assumption that fundamental information publicly available in the past has some predictive relationships to the future stock returns

1. INTRODUCTION

Today we live and breathe data. Forecasting the stock exchange data is an important financial subject which involves an assumption that the fundamental information publicly available in the past has some predictive relationships to the future stock returns. Stock market forecasting contains uncovering the market trends, planning investment tactics, identifying the best time to purchase the stocks and which stocks to purchase. A stock exchange or equity business sector is a non-direct, non-parametric framework that is difficult to model with any sensible exactness. It is the mix of speculators who need to purchase or offer or hold a share at a specific time. Prediction will continue to be an exciting locale of research, making scientists in the analytics field always desiring to enhance the existing forecasting models.

The motivation is that companies and individuals are empowered to make investment decisions to develop viable systems about their future endeavors.

Stock price prediction is a heated topic in prediction study of the financial area. The stock market is essentially a non-linear, nonparametric system that is extremely hard to model with any reasonable accuracy. Investors have been trying to find a way to predict stock prices and to find the right stocks and right timing to buy or sell. Most of the techniques used in technical analysis are highly subjective in nature and have been shown not to be statistically valid. Recently, data mining techniques and artificial intelligence techniques like decision trees, rough set approach, and artificial neural networks have been applied to this area. Data mining refers to extracting or mining knowledge from large data stores or sets. Some of its functionalities are the discovery of concept or class descriptions, associations and correlations, classification, prediction, clustering, trend analysis, outlier and deviation analysis, and similarity analysis. Data classification can be done in many different methods; one of those methods is the classification by using Decision Tree. It is a graphical representation of all possible outcomes and the paths by which they may be reached.

The use of ANN in business environments has been increasing over the last few years. Excellent algorithm has been applied to predict stock price or index. Interest in neural networks has led to a considerable surge in research activities in the past decade. Artificial neural network models are based on the neural structure of the brain.The brain learns from experience and so do artificial neural networks. As a useful analytical tool, ANN is widely applied in analyzing the business data stored in databases or data warehouses. Identifying customer behavior patterns and predicting stock price are emerging areas of neural network research and its application. Most of the companies have created new methods of evaluating financial data and investment decisions. Artificial Neural Networks are being used by most companies for improved forecasting capabilities in analysis of stock market. So, artificial neural networks suits better than other models in predicting the stock market.

The idea of forecasting using a neural network is to find an approximation of mapping between the input and output data through training. The trained neural network is then used to predict the values for the future This research work presents the use of artificial neural network as a forecasting tool for predicting the stock market price.

Mostly the approaches are in terms of fundamental approach and technical approach. For the long-term valuation fundamental approach is used. Every stock is having its own value that does not depend on the price of the stock that is known as Intrinsic value. The proposed model works through phases of data collection, feature processing, fuzzy logic mapping and stock value calculation. Fuzzy logic is used to map the quality as well as quantity valuation factors. The IF THEN rules are applied on the linguistic variable. The fuzzy model outcomes the stock value which is used to provide stock worth. The stock value is calculated by the Dividend discount model. Accuracy of the system is 0.77. The results offer the backbone for the value and not the price.

Another method is DATA MINING. Decision making process for business can be risky. Corporate decision makers have to make decisions to protect the company‟s benefit and lower the risk. In order to evaluate a data mining approach on forecasting, a tool, called IFF, was developed for evaluating and simulating forecasts. Specifically data mining techniques‟ and simulation‟s ability to predict, evaluate and validate Port Industry forecasts is tested. Accuracy is calculated with data mining methods. Finally the probability of the user‟s and simulation model‟s confidentiality is calculated. The results of the research indicate that data mining approach on forecasting and Monte Carlo method have the capability to forecast on Port industry and, if properly analyzed, can give accurate results for forecasts.

We study a multivariate Markov chain model for categorical data sequences to fuzzy time series. The proposed method gets a higher average forecasting accuracy rate than some of the existing methods on temperature prediction.

The future stock returns have some predictive relationships with the publicly available information of present and historical stock market indices. ARIMA is a statistical model which is known to be efficient for time series forecasting especially for short-term prediction. In this paper, we propose a model for forecasting the stock market trends based on the technical analysis using historical stock market data and ARIMA model. This model will automate the process of direction of future stock price indices and provides assistance for financial specialists to choose the better timing for purchasing and/or selling of stocks. The results are shown in terms of visualizations using R programming language. The obtained results reveal that the ARIMA model has a strong potential for short-term prediction of stock market trends.

2. PURPOSE

This project aims to meet the following objectives:

i. To develop such an automated system that enables users to automate the process of stock price prediction using this deep learning model.

ii. To test the accuracy of the model

iii. Model is very effective in showing which stock is good to invest your money in.

3. MOTIVATION AND SCOPE

Here we can apply such models of machine learning which works well when we use big datasets , that it does not give errors in between . And so we have used LSTM Long Short Term Memory as a sort of intermittent neural system. In RNN yield from the last advance is taken care of as contribution to the present advance. LSTM was designed by Hochreiter and Schmidhuber. It handled the issue of long haul conditions of RNN in which the RNN can't anticipate the word put away in the drawn out memory yet can give progressively exact expectations from the ongoing data. As the whole length builds RNN doesn't give efficient

execution. LSTM can be an option ,as it holds the data for extensive stretches of time. It is utilized for handling, anticipating and characterizing based on time arrangement information.

And this works efficiently when we have to use large datasets . That it can easily deal with the large no. Of data

4. LITERATURE SURVEY

The list of stock markets authors have obtained their data for training and testing of their prospective models is shown in Table 1. Surveyed articles focus on forecasting returns of a single stock market index or of multiple stock market indexes. However, several studies concentrate on forecasting returns of a single stock or multiple stocks (Ajith, Baikunth, & Mahanti, 2003a; Atsalakis & Valavanis, 2006a, 2006b).

Articles in Table 1 may be classified in three categories.

- **The first category** includes articles that are used as input data indexes from well developed markets in Western Europe, North America and other solid economy countries.

Ettes (2000), Setnes and Van Drempt (1999) model the Amsterdam stock exchange. Brownstone (1996), Kanas and Yannopoulos (2001) model the FTSE stock index. Lendasse, De Bodt, Wertz, and Verleysen (2000) studied the Belgian market. The Madrid stock exchange is examined by Fernandez-Rodriguez, Gonzalez-Martel, and Sevilla-Rivera (2000), Perez-Rodriguez, Torrab, and Andrada-Felixa (2004). The German DAX is forecasted by Rast (1999), Schumann and Lohrbach (1993), Siekmann, Gebhardt, and Kruse (1999), Steiner and Wittkemper (1997). These markets belong to well developed European markets.

Baek and Cho (2002), Chun and Park (2005), Kim and Chun (1998), Kim (1998), Oh and Kim (2002) model the Korean stock index. Cao, Leggio, and Schniederjans (2005), Yiwen, Guizhong, and Zongping (2000), Zhang, Jiang, and Li (2004), Zhongxing and Liting (1993)

examine the Shangai stock market. Lam (2001) forecasts the Hong Kong stock exchange, Chen, Leung, and Daouk (2003), Kuo (1998), Wang (2002), Wang and Leu (1996) study the Taiwan stock index. Baba and Kozaki (1992), Huang, Nakamori, and Wang (2005), Jaruszewicz and Mandziuk (2004), Kimoto, Asakawa, Yoda, and Takeoka (1990), Mizuno, Kosaka, and Yajima (1998) forecast the Japanese Stock.

Ajith et al. (2003a), Ajith, Sajith, and Sarathchandran (2003b), Chen, Abraham, Yang, and Yang (2005a) attempt to forecast the NASDAQ stock exchange, and Chaturvedi and Chandra (2004), Halliday (2004), Leigh, Paz, and Purvis (2002) try to forecast the NYSE. The S&P 500 has the highest percentage of preference among studies as in Armano, Marchesi, and Murru (2004), Casas (2001), Malliaris and Salchenberger (1993), Tsaih, Hsu, and Lai (1998). Olson and Mossman (2003) studies the Toronto stock exchange index. Surveyed markets are the North America well developed markets. This survey includes also studies from the Australian Stock Index, surveyed by Barnes, Rimmer, and Ting (2000), Pan, Tilakaratne, and Yearwood (2005), Vanstone, Finnie, and Tan (2005).

- **The second category** focuses on studies that use indexes to forecast emerging markets. Studies from ex-Eastern Europe include Zorin and Borisov (2002) for the Latvian Riga stock exchange index, Walczak (1999), Wikowska (1995) for the Polish stock exchange index. Egeli, Ozturan, and Badur (2003), Yumlu, Gurgen, and Okay (2004, 2005) forecast the Istanbul Stock Exchange market. From Western European emerging markets, studies include Koulouriotis Koulouriotis (2004, 2001, 2002, 2005) for the Athens stock exchange, Andreou, Neocleous, Schizas, and Toumpouris (2000), Constantinou, Georgiades, Kazandjian, and Kouretas (2006) for the Cyprus stock exchange. The Singapore stock exchange is the most popular emerging market, forecasted by Ayob, Nasrudin, Omar, and Surip (2001), Hui, Yap, and Prakash (2000), Kim (1998), Phua, Hoh, Daohua, and Weiding (2001).
- **The third category** includes articles that do not focus on a particular stock exchange market index, but use independent stocks or portfolio of stocks, instead. A typical example of this category is the study by Pantazopoulos, Tsoukalas, Bourbakis, Bruen, and Houstis (1998) that uses as input the price of the IBM stock, the study of Steiner and Wittkemper (1997), who selected as inputs the 16 top/bottom stocks from

the DAX and the research by Atsalakis and Valavanis (2006a, 2006b) applied to five stocks of the Athens Stock Exchange and the NYSE.

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. 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.

5. PROBLEM STATEMENT

Fundamental and technical analyses were the first two methods used to forecast stock prices.Artificial Neural networks (ANNs) is the most commonly used technique [1]. In most cases ANNs suffer from overfitting problem due to the large number of parameters to fix, and the little prior user knowledge about the relevance of the inputs in the analyzed problem [2].

Also, Support vector machines (SVMs) had been developed as an alternative that avoids such limitations. Their practical successes can be attributed to solid theoretical foundations based on VC-theory [3]. SVM computes globally optimal solutions, unlike those obtained with ANNs, which tend to fall into local minima [4].

Least squares –support vector machines (LS-SVM) method was presented in [5], which reformulated the traditional SVM algorithm. LS-SVM uses a regularized least squares function with equality constraints, leading to a linear system which meets the Karush-Kuhn-Tucker (KKT) conditions for obtaining an optimal solution.

6. EXISTING SYSTEM

The technology method was the first method to be compatible with the basic method used to predict inflation. But then we got the Artificial neural Network (ANN) and it contained a problem in predicting the future stock price, due to the large number of variables and the limited information of private users about the similarity of the problem analysis, the overfitting became a problem. Initially we worked perfectly along with the trouble of overfitting, until that time when we got the Long Short Memory i.e LSTM.

7. PROPOSED METHODOLOGY

Here we have used LSTM Long Short Term Memory is a sort of intermittent neural system. In RNN yield from the last advance is taken care of as contribution to the present advance. LSTM was desgined by Hochreiter and Schmidhuber. It handled the issue of long haul conditions of RNN in which the RNN can't anticipate the word put away in the drawn out memory yet can give progressively exact expectations from the ongoing data. As the whole length builds RNN doesn't give efficient execution. LSTM can be an option ,as it holds the data for extensive stretch of time. It is utilized for handling, anticipating and characterizing based on time arrangement information.

8. SYSTEM DESIGN

There is a lot of similar research for stock price forecasts. Vector equipment was used to create a model to retrieve stock history data and to forecast stock trends. The Particle swarm optimization algorithm is used to add a variable vector, support vector (svm), which can predict the stock's value by force. This study improves the mechanism of the vector support system, but the synthetic particle (pso) algorithm requires longer calculations. LSTM is integrated with a naive Bayesian way of extracting market sentiment to improve forecast performance. This method can be used to predict financial markets at time scales that are completely different from other variables. The emotion analysis model combined with the LSTM series learning model for the acquisition of a dynamic stock price

forecasting model, and the results showed that this model could improve the forecast accuracy. Jia has talked about the effectiveness of LSTM in predicting stock price volatility, and research showed that LSTM is an effective way to predict stock profit. Real-time wavelet denoising was integrated with the LSTM network to predict the index of east Asian stocks, correcting some of the logical contradictions in previous studies. Compared with the original LSTM, this composite model was significantly improved with higher prediction accuracy and less repeat error. The Bagging method has been used to integrate the neural net- work method of forecasting the stock index of China (including the Shanghai index) with different accuracy of different stock index predictions, but the prediction of proximity is not satisfactory. An evolutionary approach was used to predict the trend of stock price fluctuations. A network of deep convictions about natural plastics was used to predict the time series of stocks. The neural exchange network was used to determine stock price trends. A neural network-based network model was developed for future stock price forecasts using a hybrid approach that combines variance in technical analysis with basic variations of stock market indicators with the BP algorithm. The results indicate that this method has greater accuracy in predicting daily stock price than the technical analysis method. The soft computer technology(SCT), which was developed for the Dhaka Stock Exchange (DSE) to foresee the closing price of the DSE. The feeling of comparing the artificial neural network with the neural fuzzy consultation system suggests that this approach works well.

The Artificial bee colony algorithm is integrated with wave- let changes and a continuous neural network for stock price prediction. A number of international stock indexes were developed for evaluation, including the Dow Jones industrial index (DJIA), the London FTSE 100 index (FTSE), the Tokyo Nikkei-225 (Nikkei) index and the Taiwan stock agency Capitalization Weighted stock Index (TAIEX). The simulation results show that the system performs well in forecasting and can be used in real-time stock trading trading systems.

Multiple output speaker models based on RNN-LSTM have been used in the field of speech recognition. Experimental results show that the model is better than the single speaker model, as well as the infrastructure under which new branches are installed.

Finding a new exit model not only reduces memory usage but is also better than training a new platform model. Model of a multi-input neural input network model. (MIMO-Net) was designed for cell division of microscope images. Experimental results show that this approach is more than just a high-quality, youth-based learning method.

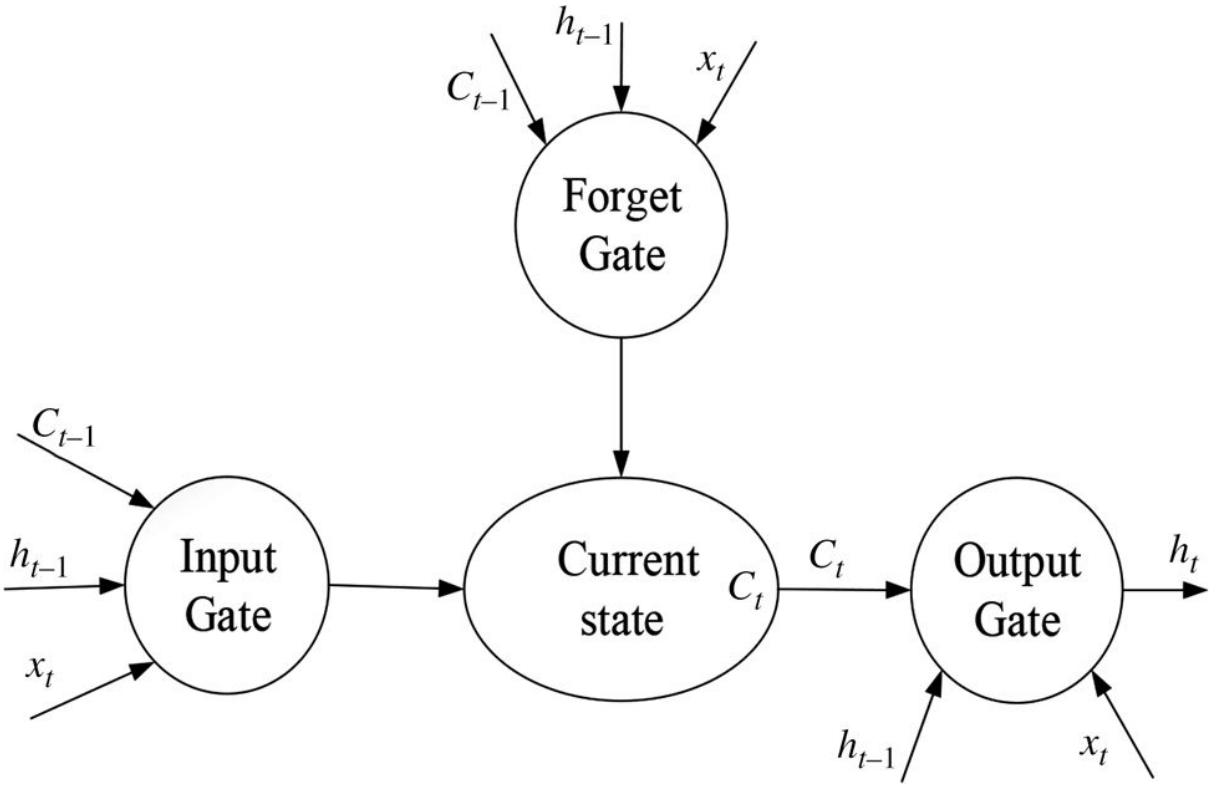
Inspired by the above research, considering that other parameters and stock indicators combined, it is necessary to design

9. LIST OF FIGURES

Long Short Term Memory Networks (LSTM) are a developmental approach to standard neural networks (RNN), which makes it easy to recall past data in memory. The disappearing RNN problem is solved here. Short-Term Memory (LSTM) is well suited to classifying and predicting time series given an unknown downtime. It trains the model through back propagation. As given in fig below. In the LSTM network, there are three gates.

It has three “gate” buildings (shown in Where). The three gates are designated by the LSTM unit, known as the input gate, the output gate and the output gate. While data is entered into the LSTM network, it can be selected by rules. Only the details that match the algorithm will be omitted, and the incompatible information will be remembered by the forgotten gateway.

The gateway allows information to be transmitted selectively with Eq. 1 shows the default implementation function of the LSTM net, the sigmoid function. LSTM can add and delete information of neurons per gating unit. Deciding on whether or not the information is passing.



➢ **Input gate** — find which esteem from information ought to be utilized to alter the memory. Sigmoid capacity chooses which esteems to let through 0,1. furthermore, tanh work offers weightage to the qualities which are passed choosing their degree of significance going from-1 to 1.

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

➢ **Forget gate** — find what subtleties to be disposed of from the square. It is chosen by the sigmoid capacity. it takes a gander at the past state(ht-1) and

the substance input(Xt) and yields a number between 0(omit this)and 1(keep this)for each number in the cell state Ct−1.

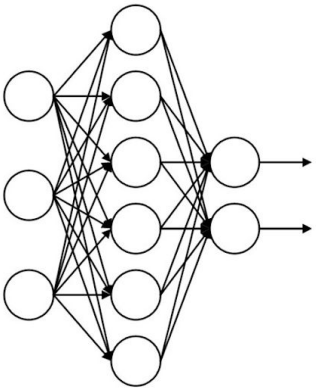
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

- **Output gate** —The square data and memory are used to select the crop. The Sigmoid slope determines which equations allow approval 0,1. In addition, the tanh function gives weight to the passing attributes choosing their size from 1 to 1 and doubling the Sigmoid fruit.

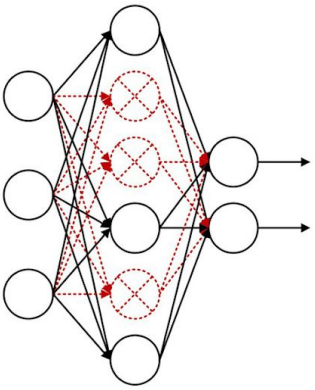
$$o_t = \sigma\left(W_o \left[h_{t-1}, x_t\right] \; + \; b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

-
- **DEEP RECURRENT NEURAL NETWORK**



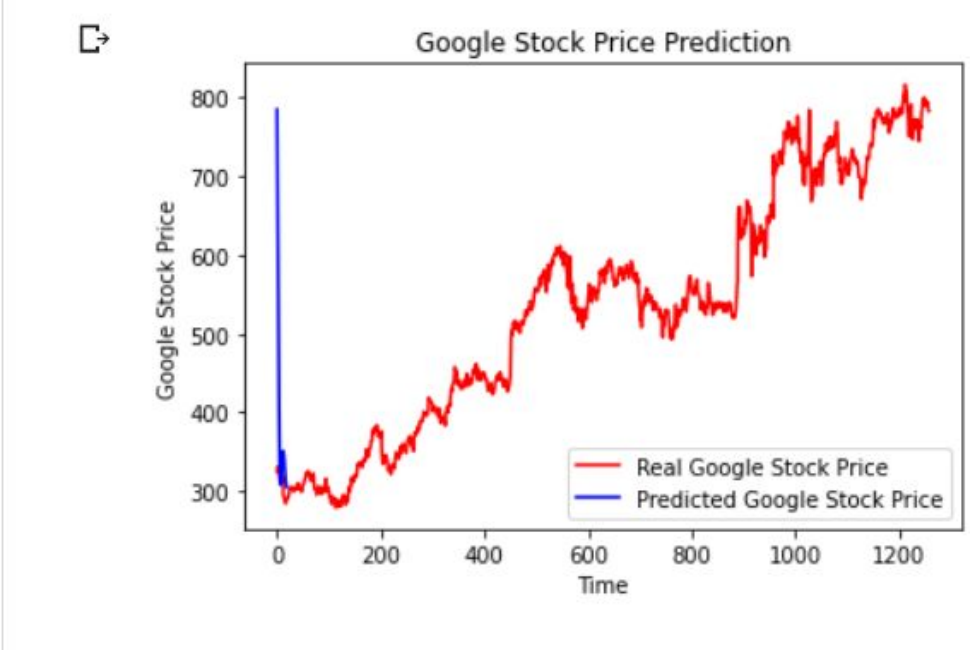
(a) The neural network model without dropout



(b) The neural network model with dropout

10. RESULT

➢ **Google Stock Prediction Graph**



11. CONCLUSION AND FUTURE SCOPE

The investment that we make in stock is always one of the best channels for cash flow. However possible, it is always difficult to choose the best time to buy or sell because of the volatile and volatile nature of the securities exchange. The way to receive maximum returns on the stock exchange is to determine the right exchange period when the exchange risk should be minimal. Since stock exchange is a risky business, it is important to read the stock records, their developments and to assess the risks and benefits before going on an exchange. In this figure the three important components of time-series assessments such as, stock price expectations, volatility and stock price volatility, following stock valuation are illustrated using machine insight and variable registry methods. Term Memory (LSTM), which is a type of RNN, will have the option to choose when the curator can withdraw its money. In addition, there will be greater profitability, Secure Securities trading is one of the most common forms of cash-related activities. However, it is difficult to choose the best time to buy or sell because of the volatile and fluctuating earnings and relative volatility of cash-related trading. A good way to see the maximum benefit in stock trading is to choose a reasonable trading period when the risk of trading should be minimal. Since stock trading is a risky business, it is important to foresee stock records, its turnaround events and risk assessments and focal points before committing to any trading. In this proposal large pieces of capital related to the timing of the performance appraisal, for example, a fair desire for stock, demonstration and evaluation of the stock and trade following a request for improvement of the already estimated stock are shown using machine information and structured plans.

Long Short-term memory is widely used here, because it remembers the simplest background data from all the other tools available, such as the Artificial Neural Network and the support vector machine. So after seeing all the records of machine learning algorithms we came to a conclusion to use LSTM for predicting the value of stock in future

12. REFERENCES

1. Gupta A. :"*Stock market prediction by help of Hidden Markov Models*",, 2012 Students Conference on, pp.1-4, 2012, IEEE Engineering and Systems (SCES)
2. Jia H *"Investigation into the effectiveness of the long short term memory networks for stock price prediction".*
3. Kaggle is the source for *Stock Price Prediction* dataset.
4. K Hiba Sadia,, Adarrsh Paul, Aditya Sharma, Saurav Sanyal "*Stock Market Prediction Using Machine Learning Algorithms*"
5. Lipo W., Shekhar G.: "*Neural Networks and Wavelet De-Noising for Stock Trading and Prediction*", Springer, Time Series Analysis, Modeling and Applications Intelligent
6. Omar S. Soliman and Mustafa Abdul Salam ,Osman Hegazy "*A Machine Learning Model for Stock Market Prediction"*

.