# A Project Report

## on

# STOCK PRICE PREDICTION USING MACHINE LEARNING

*Submitted in partial fulfillment of the*

*requirement for the award of the degree*

*of*

# Bachelor of Technology in Computer Science and Engineering



**Under The Supervision of**

**Mr. K.M.BALAMURUGAN**

**Assistant Professor**
**Department of Computer Science and Engineering**

**Submitted By**

19SCSE1010095 – KAKUL NIGAM

**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING**

**DEPARTMENT OF COMPUTER SCIENCE AND**

**ENGINEERING GALGOTIAS UNIVERSITY, GREATER NOIDA,**

**INDIA DECEMBER - 2021**

# SCHOOL OF COMPUTING SCIENCE
# AND ENGINEERING
# GALGOTIAS UNIVERSITY, GREATER NOIDA

## CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the project, entitled **" Stock Price Prediction Using Machine Learning"** in partial fulfillment of the requirements for the award of the **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING** submitted in the **School of Computing Science and Engineering** of Galgotias University, Greater Noida, is an original work carried out during the period of **JULY-2021 to DECEMBER-2021**, under the supervision of **Mr.K.M BALAMURUGAN, Assistant Professor**, **Department of Computer Science and Engineering** of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the project has not been submitted by me for the award of any other degree of this or any other places.

19SCSE1010095 – KAKUL NIGAM

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor

DR.K.M. BALAMURUGAN

(ASSISTANT PROFESSOR)

## CERTIFICATE

The Final Project Viva-Voce examination of **19SCSE1010095 – KAKUL NIGAM,** has been held

on _24/12/2021 _and his/her work is recommended for the award of **BACHELOR OF**

**TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING**.

**Signature of Examiner(s)**                                     **Signature of Supervisor(s)**

**Signature of Dean**

Date:

Place:

# ABSTRACT

The covid -19 pandemic has not only been a traumatic experience for the whole world healthwise, but it also has been a disastrous hindrance for the economical growth of the world. Number of companies and organizations went bankrupt or were shut down, millions of people lost their jobs and there was a drastic change in the stock prices of some of the biggest organizations in the world during the prime of the covid-19 pandemic. No one could have predicted these shocking changes in the stock prices as no one saw it coming. Since there was a drastic change in the stock prices so now there is a lot of new important data that needs to be used for the further prediction of the stock prices in the future. In this paper we will predict the stock prices using the new data after the prime of the covid-19 pandemic using machine learning algorithms.

# Table of Contents

# List of Figures

## Acronyms

| SVM | Support Vector Maching |
|-----|------------------------|
| ML  | Machine Learning |
| DL  | Deep Learning |
| CNN | Convolution Neural Networks |
|     |  |
|     |  |
|     |  |

# CHAPTER-1

## Introduction

There are a lot of complicated financial indicators and also the fluctuation of the stock market is highly violent. However, as the technology is getting advanced, the opportunity to gain a steady fortune from the stock market is increased and it also helps experts to find out the most informative indicators to make a better prediction. The prediction of the market value is of great importance to help in maximizing the profit of stock option purchase while keeping the risk low. Recurrent neural networks (RNN) have proved one of the most powerful models for processing sequential data. Various types of neural networks can be developed by the combination of different factors like network topology, training method etc. For this experiment, we have considered Recurrent Neural Network and Long Short-Term Memory.

# CHAPTER-2

## Literature survey

> ➢ Importance of stock market**:**

The Stock is essentially a share in a company's ownership. Stocks are partial ownership of businesses instead of stock tickers piece of paper, which can be traded in the stock market . If company ownership is divided into 100 parts, the investor purchase one part which is equal to one share then we can own 1 percent of that company. Stock exchange uses an automated matching system driven by order. Stock prices are defined as any time how many buyers and sellers available for the same stock in the market. If the number of buyers is more than sellers then stock price becomes high and if the number of sellers higher than buyers then stock price becomes low. The best buy and sell order are looked into a counterparty angle. The best buy order is which has the highest price and best sell order is which has the lowest price. With this logic system can match the orders and executes the traders'system. SEBI (Security and Exchange Board of India) regulates the stock market. In stock markets customers preferences and requirements are different. The estimated world stock market was at $36.6 trillion in early October 2008 [6]. The total world market for derivatives was estimated at approximately $791 trillion in face value or nominal value, 11 times thesize of the world economy.

➢ Prediction Techniques:

Presented the recent methods for the prediction of stock market and give a comparative analysis of all these Techniques. Major prediction techniques such as data mining, machine learning and deep learning techniques used to estimate the future stock prices based on these techniques and discussed their advantages and disadvantages. They are,

1. Holt-Winters
2. Artificial Neural Network
3. Hidden Markov Model
4. ARIMA Model
5. Time Series Linear Model
6. Recurrent Neural Networks.

Holt-Winters, Artificial Neural Network, Hidden Markov Model are Machine Learning Techniques, ARIMA Model is Time series technique and Time series Linear Model and Recurrent Neural Networks are Deep Learning Techniques.

➢ Recurrent Neural Network:

Recurrent neural networks (RNN) [17] use back propagation to learn, but their nodes have a feedback mechanism. Because of this, RNN models can predict a stock price based on recent history and is recurrent.

```
model <-  trainr(Y = trainy,  X = trainx,  learningrate = 0.05,  hidden_dim = 5,    numepochs
        = 2600)
pred <-  predictr(model,testx).
```

Advantage :

Previous time points to input layer contains inputs.

Disadvantage:

An RNN makes it possible to feed those words in through a much smaller set of input nodes.

The popularity of stock market trading is growing rapidly, which is encouraging researchers to find out new methods for the prediction using new techniques. The forecasting technique is not only helping the researchers but it also helps investors and any person dealing with the stock market. In order to help predict the stock indices, a forecasting model with good accuracy is required.

# Chapter-3

## Product design

➢ The prediction methods can be roughly divided into two categories, statistical methods and artificial intelligence methods. Statistical methods include logistic regression model, ARCH model, etc. Artificial intelligence methods include multi-layer perceptron, convolutional neural network, naive Bayes network, back propagation network, single-layer LSTM, support vector machine, recurrent neural network, etc. They used Long short-term memory network (LSTM).

➢ This section we will discuss the product design of our system. Our system consists of several stages which are as follows:

1. Stage 1: Raw Data

2. Stage 2: Data Preprocessing

3. Stage 3: Feature Extraction

4. Stage 4: Training Neural Network

5. Stage 5: Output Generation

➢ **Use Case Diagram:**

In the Unified Modelling Language (UML), a use case diagram can summarize the details of your system's users (also known as actors) and their interactions with the system. To build one, you'll use a set of specialized symbols and connectors. An effective use case diagram can help your team discuss and represent:

- Scenarios in which your system or application interacts with people, organizations, or external systems.

- Goals that your system or application helps those entities (known as actors) achieve.
- The scope of your system.
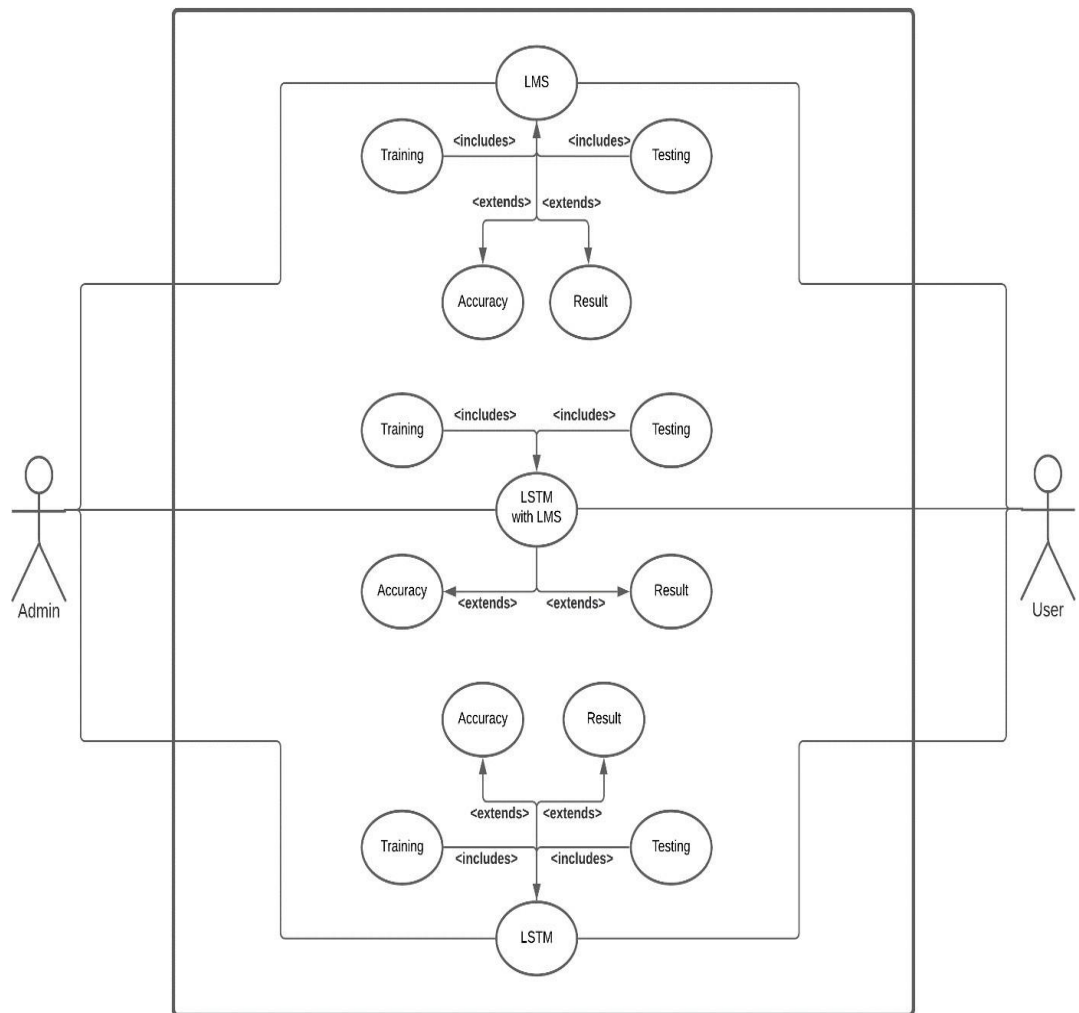


Using LMS, LSTM and LSTM with LMS in the syst

- ➤ **Sequence Diagram**

        A sequence diagram is a type of interaction diagram because it describes how and in what order a group of objects works together. These diagrams are used by software developers and business professionals to understand requirements for a new system or to document an existing process. Sequence diagrams are sometimes known as event diagrams or event scenarios. Sequence diagrams can be useful references for businesses and other organizations. Try drawing a sequence diagram to:

- Represent the details of a UML use case.
- Model the logic of a sophisticated procedure, function, or operation.
- See how objects and components interact with each other to complete a process.
- Plan and understand the detailed functionality of an existing or future scenario.
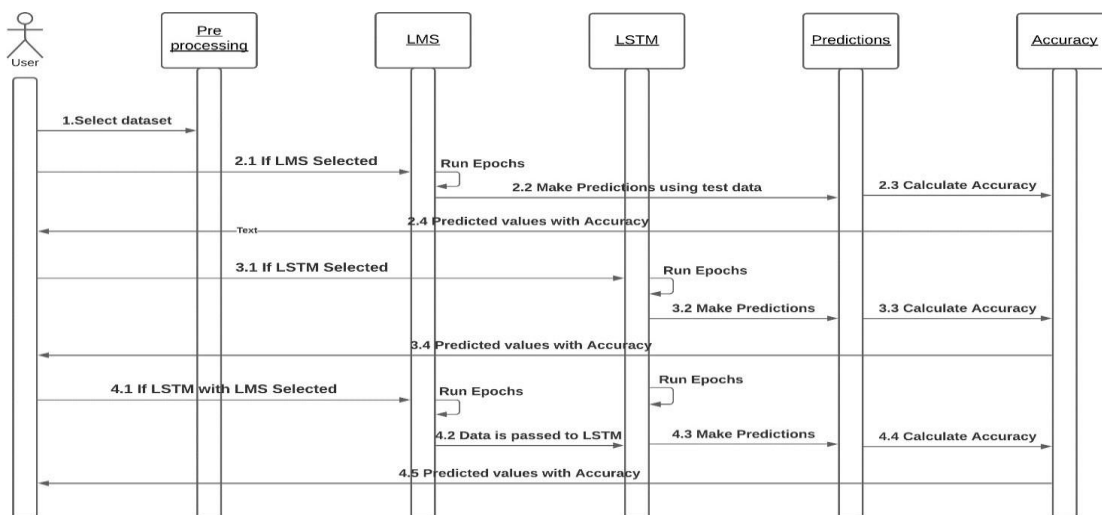


Execution based on model selection

➢ **Activity Diagram**

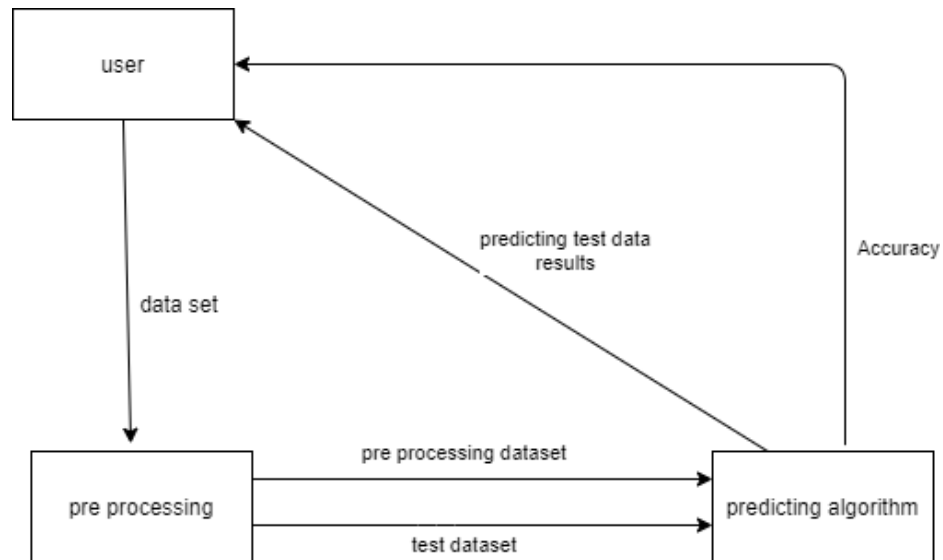An activity diagram is a behavioral diagram i.e. it depicts the behavior of a system.

An activity diagram portrays the control flow from a start point to a finish

point showing the various decision paths that exist while the activity is being

executed.
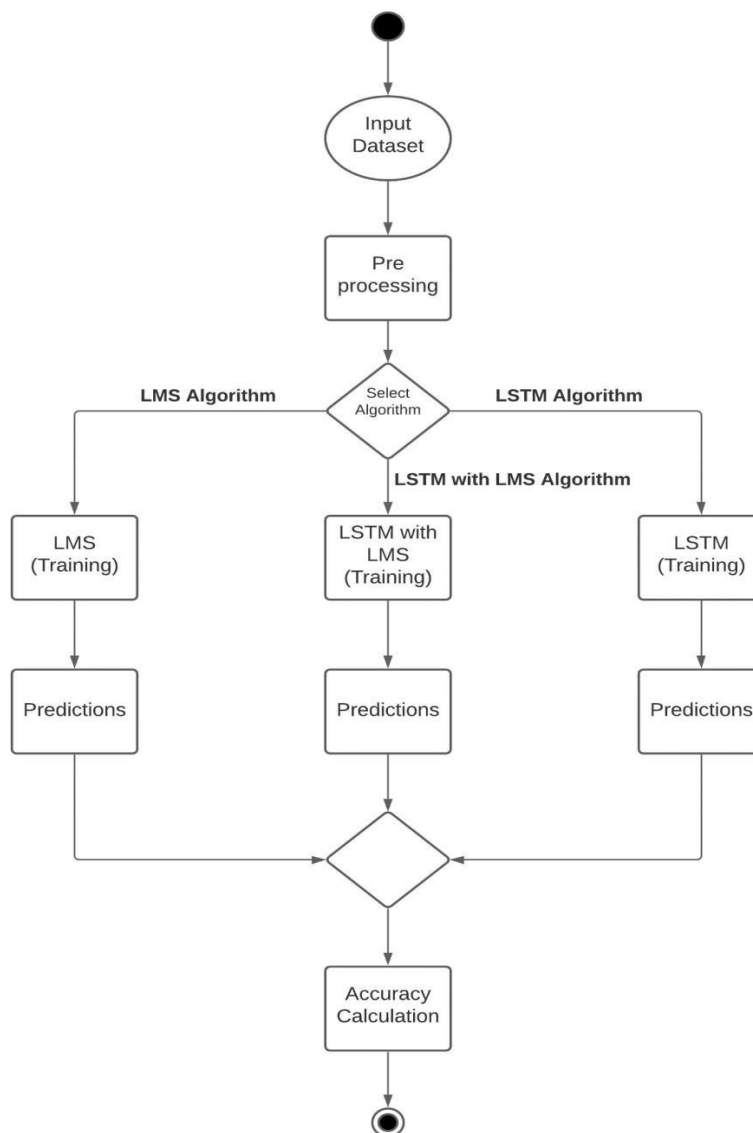


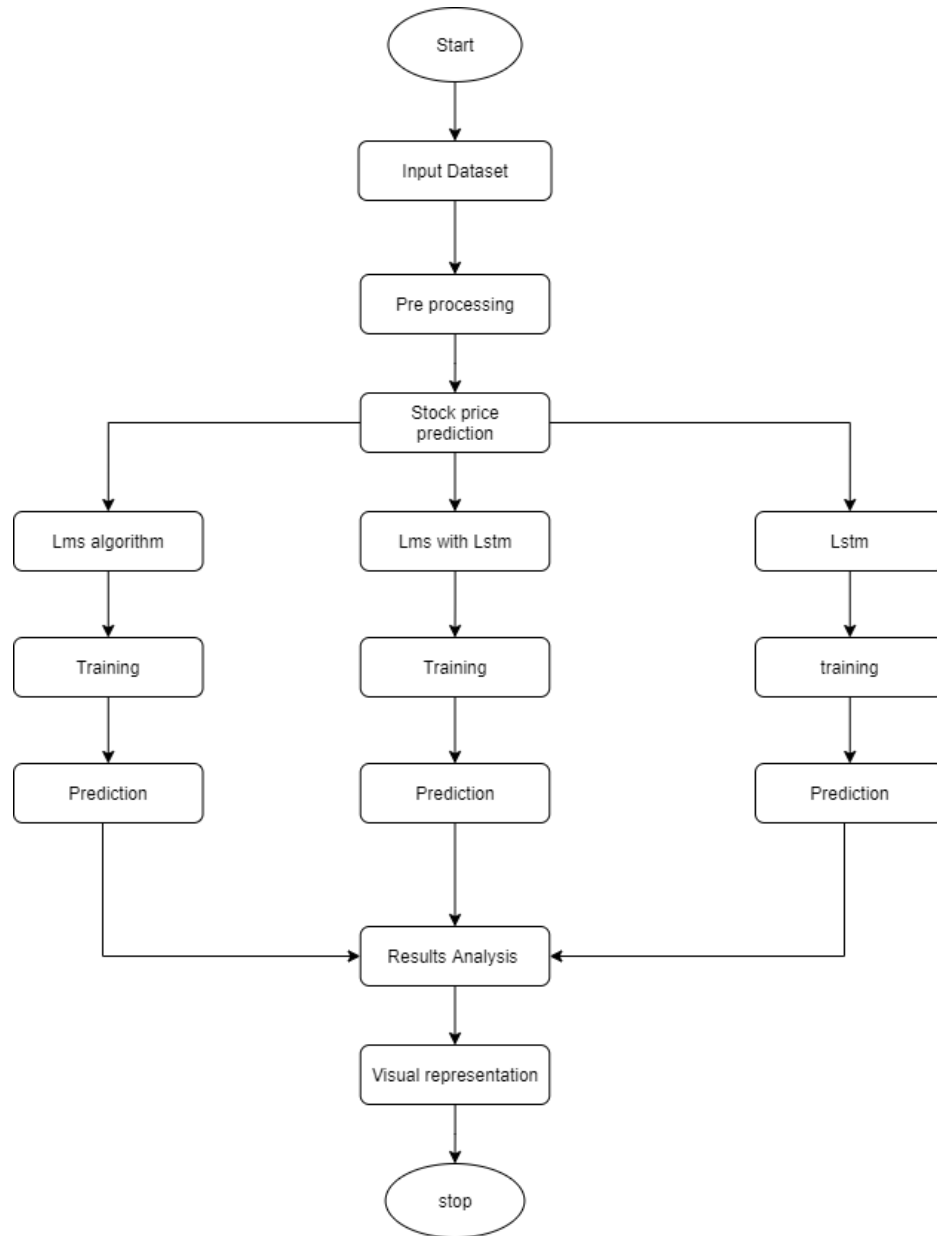Execution based on algorithm selection

## ➢ Collaboration Diagram:

The collaborations are used when it is essential to depict the relationship between the object. Both the sequence and collaboration diagrams represent the same information, but the way of portraying it quite different. The collaboration diagrams are best suited for analyzing use cases.
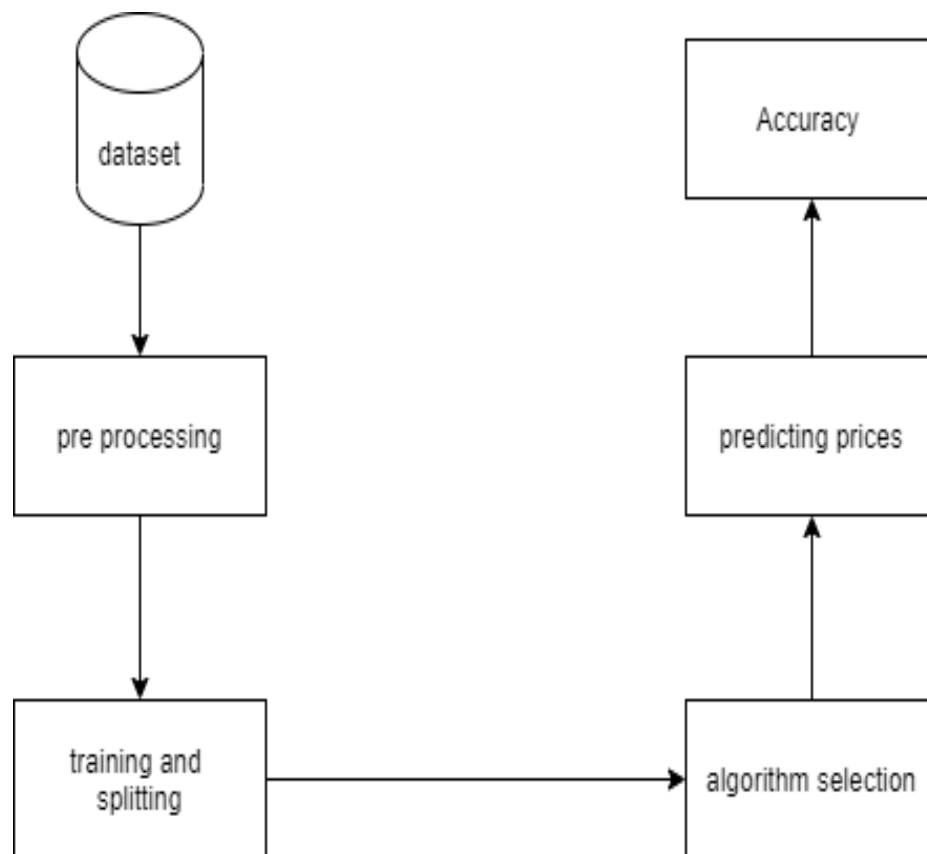
Data transfer between modules

## ➢ **Flow Chart:**



Flow of execution

## ➢ Component Diagram



dataset

Accuracy

pre processing

predicting prices

training and splitting

algorithm selection

Components present in the system

# Chapter-4

# Working

LSTM is a special network structure with three "gate" structures. Three gates are placed in an LSTM unit, called input gate, forgetting gate and output gate. While information enters the LSTM's network, it can be selected by rules. Only the information conforms to the algorithm will be left, and the information that does not conform will be forgotten through the forgetting gate. The experimental data in this paper are the actual historical data downloaded from the Internet. Three data sets were used in the experiments. It is needed to find an optimization algorithm that requires less resources and has faster convergence speed.

• Used Long Short-term Memory (LSTM) with embedded layer and the LSTM neural network with automatic encoder.

• LSTM is used instead of RNN to avoid exploding and vanishing gradients.

• In this project python is used to train the model, MATLAB is used to reduce dimensions of the input. MySQL is used as a dataset to store and retrieve data.

• The historical stock data table contains the information of opening price, the highest price, lowest price, closing price, transaction date, volume and so on.

• The accuracy of this LSTM model used in this project is 57%.

The LMS filter is a kind of deposed for solving linear problems. The idea of the filter is to minimize a system (finding the filter coefficients) by minimizing the least mean square of the error signal.
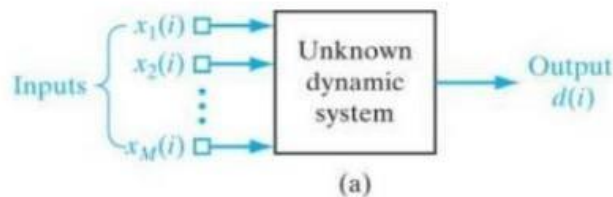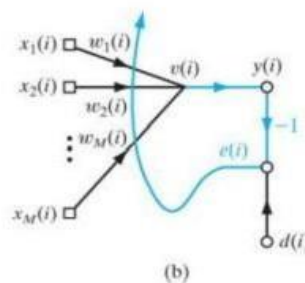


Fig. 1: LMS Inputs and Outputs          Fig 2: LMS updating weights

In general, we don't know exactly if the problem can be solved non-linear algorithm. Since the internet always shows non-linear approaches, we will use LMS to prove that stock market prediction can be done with linear algorithms with a good precision. But this filter mimetites a system, that is, if we apply this filter in our data, we will have the filter coefficients trained, and when we input a new vector, our filter coefficients will output a response that the original system would (in the best case). So we just have to do a tricky modification for using this filter to predict data.

# Chapter-5

# Implementation
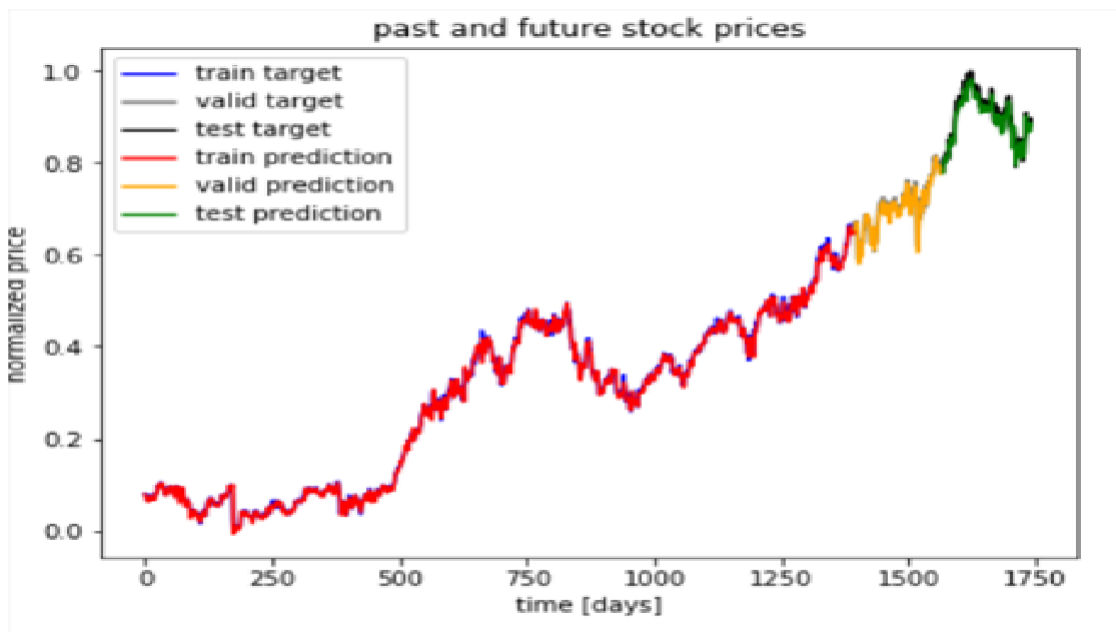


In this graph the Red line depicts train Predicition and the blue line is train target , it is observe that both the lines are almost going on the same direction and having similar features thus our train prediction is almost accurate.

Yellow line depicts the valid prediction.

Now , Green line depicts the test prediction and black colour depicts the test target , it is observe that both the lines are almost going on the  same direction and having similar featutes thus our test prediction is almost accurate.

future stock prices

In future stock prices Green colour depicts test prediction and black colour depicts the test target which show us that there is very less difference between test target and test prediction.

# REFERENCES

[1] RakhiMahant, TrilokNathPandey, Alok Kumar Jagadev, and SatchidanandaDehuri ―Optimized Radial Basis Functional Neural Network for Stock Index Prediction,‖ *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016.*

[2] Kai Chen, Yi Zhou and FangyanDai ―A LSTM-based method for stock returns prediction: A case study of China stock market,‖ *IEEE International Conference on Big Data,*2015.

[3] A.U.S.S Pradeep, SorenGoyal, J. A. Bloom, I. J. Cox, and M. Miller, ―Detection of statistical arbitrage using machine learning techniques in Indian Stock market,‖ *IIT Kanpur*, April 15, 2013.

[4] Prashant S. Chavan, Prof. Dr. Shrishail. T. Patil ―Parameters for Stock Market Prediction,‖ *Prashant S Chavan et al, Int.J.Computer Technology & Applications, Vol 4 (2),337-340.*

[5] NeelimaBudhani, Dr. C. K. Jha, Sandeep K. Budhani―Prediction of Stock Market Using Artificial Neural Network,‖ *International Conference on Soft Computing Techniques for Engineering and Technology (ICSCTET)- 2014.*

[6] SharvilKatariya, Saurabh Jain―Stock Price Trend Forecasting using Supervised Learning Methods.

[7] A. Ben-Hur and J. Weston. A user's guide to support vector machines. In O. Carugo and F. Eisenhaber (eds.), *Data Mining Techniques for the Life Sciences*, pp. 223--239. Springer, 2010.

[8] Bing Search API. Windows Azure Marketplace. http://datamarket.azure.com/dataset/bing/search

[9] N. K. Chowdhury and C. K.-S. Leung. Improved travel time prediction algorithms for intelligent transportation systems. In *Proc. KES 2011, Part II*, pp. 355—365

[10] A. Cuzzocrea, C. K.-S. Leung, and R. K. MacKinnon. Mining constrained frequent itemsets from distributed uncertain data. *FGCS*, 37: 117--126, July 2014.

[11] T. Finley and T. Joachims. Training structural SVMs when exact inference is intractable. In *Proc. ICML 2008*, ACM, pp. 304--311. ACM.

[12] D. Hodges. Is your fund manager beating the index? *MoneySense*, 15(7):10, Dec. 2013/Jan. 2014.

[13] T. Joachims, T. Finley, and C.-N. J. Yu. Cutting-plane training of structural SVMs. *Machine Learning*, 77(1): 27--59, Oct. 2009.

[14] D. E. King. Dlib-ml: a machine learning toolkit. *JMLR*, 10: 1755--1758, July 2009.

[15] V. Kolmogorov and C. Rother. Minimizing nonsubmodular functions with graph cuts-a review. *IEEE TPAMI*, 29(7): 1274--1279, July 2007.

[16] C. K.-S. Leung, F. Jiang, and Y. Hayduk. A landmark-model based system for mining frequent patterns from uncertain data streams. In *Proc. IDEAS 2011*, pp. 249--250. ACM.

[17]    C. K.-S. Leung, F. Jiang, L. Sun, and Y. Wang. A constrained frequent pattern mining system for handling aggregate constraints. In *Proc. IDEAS 2012*, pp. 14--23. ACM.

[18]    C. K.-S. Leung, S. K. Tanbeer, B. P. Budhia, and L. C. Zacharias. Mining probabilistic datasets vertically. In *Proc. IDEAS 2012*, pp. 199--204. ACM.

[19]    T. Mitchell. *Machine Learning*. McGraw Hill, 1997.

[20]    K. P. Murphy. *Machine Learning: a Probabilistic Perspective*. MIT Press, 2012.

[21]    Quandl Stock Index. http://www.quandl.com/stocks

[22]    S&P 500 information technology. S&P Dow Jones Indices. http://ca.spindices.com/indices/equity/sp-500-information-technology-sector

[23]    C. Sammut and G. I. Webb (eds.) *Encyclopedia of Machine Learning*. Springer, 2010.

[24]    S. K. Tanbeer, C. K.-S. Leung, and J. J. Cameron. Interactive mining of strong friends from social networks and its applications in e-commerce. *JOCEC*, 24(2-3): 157--173, 2014

[25]    I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *JMLR*, 6: 1453--1484, Sept. 2005