##### INDEX PRICE PREDICTION

A PROJECT REPORT

###### *Submitted by*

##### SAKSHAM MATHUR (21BCE10666)

##### AYUSH GUPTA (21BCE10360)

##### ANANYA MISHRA (21BCE10234)

##### IPSHITA DE(21BCE10899)

##### RITOJA DEY(21BCE10687)

*in partial fulfilment for the award of the degree*

*of*

##### BACHELOR OF TECHNOLOGY

*in*

# COMPUTER SCIENCE AND ENGINEERING



SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

VIT BHOPAL UNIVERSITY

KOTHRIKALAN, SEHORE

MADHYA PRADESH - 466114

SEPTEMBER 2022

VIT BHOPAL UNIVERSITY, KOTHRIKALAN, SEHORE

MADHYA PRADESH – 466114

BONAFIDE CERTIFICATE

Certified that this project report titled “INDEX PRICE PREDICTION” is the bonafide work of “SAKSHAM MATHUR (21BCE10666), AYUSH GUPTA (21BCE10360), ANANYA MISHRA (21BCE10234), IPSHITA DE (21BCE10899), RITOJA DEY (21BCE10687) ” who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported here does not form part of any other project / research work on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROGRAM CHAIR PROJECT GUIDE

Dr. PREETAM SUMAN, Dr. SANDEEP SAHU

Program Chair, CSE-Core Assistant Professor

School of Computer Science and Engineering School of Computer Science and Engineering

VIT BHOPAL UNIVERSITY VIT BHOPAL UNIVERSITY

The Project Exhibition I Examination is held on 30/09/2022

ACKNOWLEDGEMENT

First and foremost I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to Dr Preetam Suman, Head of the Department, School of Aeronautical Science for much of his valuable support encouragement in carrying out this work.

I would like to thank my internal guide Mr. Sandeep Sahu, for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work.

I would like to thank all the technical and teaching staff of the School of Computer Science and Engineering, who extended directly or indirectly all support.

Last, but not the least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

ABSTRACT

We are working on website which helps users to get predicted stock market price values for investment and trading purpose. our website is made with complete userfriendly interface.

In this project we attempt to implement machine learning approach to predict stock prices. Machine learning is effectively implemented in forecasting stock prices. The objective is to predict the stock prices in order to make more informed and accurate investment decisions. We propose a stock price prediction system that integrates mathematical functions, machine learning, and other external factors for the purpose of achieving better stock prediction accuracy and issuing profitable trades.

There are two types of stocks. You may know of intraday trading by the commonly used term "day trading." Inter-day traders hold securities positions from at least one day to the next and often for several days to weeks or months. LSTMs are very powerful in sequence prediction problems because they’re able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price. While predicting the actual price of a stock is an uphill climb, we can build a model that will predict whether the price will go up or down.

Keywords: LSTM, CNN, ML, DL, Trade Open, Trade Close, Trade Low, Trade High, Web dev, MySQL, phpMyAdmin

TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | List of Abbreviations  List of Figures  Abstract | iii  iv  v  vi |
| 1 | **INTRODUCTION** Introduction1.1.1 Time-series Prediction 1.2 Motivation for the work  1.3 Problem Statement | 1  .  .  . |
| 2 | **LITERATURE SURVEY**  2.1 Introduction  2.2 Existing Methods  2.2.1 Stock Market Prediction using ML  2.2.2 Forecasting the Stock Market Index  2.2.3 Indian Stock Market Prediction  2.2.4 The Stock Market and Investment  2.2.5 Automated Stock Price Prediction  2.2.6 LSTM Hybrid Model  2.2.7 LSTM and Random Forest |  |
| 3 | **SYSTEM ANALYSIS**  3.1 Proposed System  3.1.1 Proposed Work  3.1.2 Preprocessing of Data  3.1.3 Overall Architecture |  |
| 4 | **SYSTEM DESIGN AND IMPLEMENTATION**  4.1 Introduction  4.1.1 Working of LSTM  4.2 Hardware Requirements  4.3 Software Requirements  4.4 DataSet Details  4.5 Tools & Technologies  4.5.1 Python  4.5.2 Numpy  4.5.3 SCIKIT Learn  4.5.4 Tensorflow  4.5.5 Keras  4.6 Implementation  4.6.1 Read Dataset  4.6.2 Graph of Close History  4.6.3 Model Fitting Summaary  4.6.4 Prediction |  |
| 5 | **PERFORMANCE ANALYSIS**  5.1System Configuration  5.1.1 Hardware Requirements  5.1.2 Software Requirements  5.2 LSTM NIFTY50 (performance Analysis |  |
| 6 | **FUTURE ENHANCEMENT AND CONCLUSION**  6.1 Limitation/Constraints of the System  6.2 Conclusion  6.3 Future Enhancements |  |
|  | References |  |

LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **TITLE** | **PAGE NO.** |
| 1.1 | Work flow | 25 |
| 1.2 | Data Preprocessing | 26 |
| 1.3 | Overall Architecture | 27 |
| 2.1 | Dataset | 32 |
| 2.2 | Close History Graph | 33 |
| 2.3 | Model Summary | 33 |
| 2.4 | Prediction | 34 |
| 3 | LSTM Performance |  |

LIST OF ABBREVIATIONS

LSTM Long Short-Term Memory

ATS Automated Trading System

GRU Gated Recurrent Unit

ML Machine Learning

SVM Support Vector Machine

EMH Efficient Market hypothesis

AI Artificial Intelligence

NN Neural Networks

ARMA Autoregressive Moving Average

DRL Deep Reinforcement Learning

LMS Least Mean Square

UML Unified modelling Language

MSE Mean Squared Error

RMSE Root Mean Squared Error

**CHAPTER 1**

**1. INTRODUCTION**

# The financial market is a dynamic and composite system where people can buy and sell currencies, stocks, equities and derivatives over virtual platforms supported by brokers. The stock market allows investors to own shares of public companies through trading either by exchange or over the counter markets. This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money, low risk compared to the risk of opening new business or the need of high salary career. Stock markets are affected by many factors causing the uncertainty and high volatility in the market. Although humans can take orders and submit them to the market, automated trading systems (ATS) that are operated by the implementation of computer programs can perform better and with higher momentum in submitting orders than any human. However, to evaluate and control the performance of ATSs, the implementation of risk strategies and safety measures applied based on human judgements are required. Many factors are incorporated and considered when developing an ATS, for instance, trading strategy to be adopted, complex mathematical functions that reflect the state of a specific stock, machine learning algorithms that enable the prediction of the future stock value, and specific news related to the stock being analysed.

**1.1.1 Time-series Prediction:**

Time-series prediction is a common technique widely used in many real-world applications such as weather forecasting and financial market prediction. It uses the continuous data in a period of time to predict the result in the next time unit. Many time- series prediction algorithms have shown their effectiveness in practice. The most common algorithms now are based on Recurrent Neural Networks (RNN), as well as its special type - Long-short Term Memory (LSTM) and Gated Recurrent Unit (GRU). Stock market is a typical area that presents time-series data and many researchers study on it and proposed various models. In this project, LSTM model is used to predict the stock price.

**1.2 MOTIVATION FOR WORK**

# Businesses primarily run over customer’s satisfaction, customer reviews about their products. Shifts in sentiment on social media have been shown to correlate with shifts in stock markets. Identifying customer grievances thereby resolving them leads to customer satisfaction as well as trustworthiness of an organisation. Hence there is a necessity of an un biased automated system to classify customer reviews regarding any problem. In today’s environment where we’re justifiably suffering from data overload (although this does not mean better or deeper insights), companies might have mountains of customer feedback collected; but for mere humans, it’s still impossible to analyse it manually without any sort of error or bias. Oftentimes, companies with the best intentions find themselves in an insights vacuum. You know you need insights to inform your decision making and you know that you’re lacking them, but don’t know how best to get them. Sentiment analysis provides some answers into what the most important issues are, from the perspective of customers, at least. Because sentiment analysis can be automated, decisions can be made based on a significant amount of data rather than plain intuition.

**1.3 PROBLEM STATEMENT**

Time Series forecasting & modelling plays an important role in data analysis. Time series analysis is a specialized branch of statistics used extensively in fields such as Econometrics & Operation Research. Time Series is being widely used in analytic & data science. Stock prices are volatile in nature and price depends on various factors. The main aim of this project is to predict stock prices using Long short term memory (LSTM).

**Chapter 2**

# LITERATURE SURVEY

**2.1 INTRODUCTION**

# "What other people think” has always been an important piece of information for most of us during the decision-making process. The Internet and the Web have now (among other things) made it possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics — that is, people we have never heard of. And conversely, more and more people are making their opinions available to strangers via the Internet. The interest that individual users show in online opinions about products and services, and the potential influence such opinions wield, is something that is driving force for this area of interest. And there are many challenges involved in this process which needs to be walked all over in order to attain proper outcomes out of them. In this survey we analysed basic methodology that usually happens in this process and measures that are to be taken to overcome the challenges being faced.

**2.2EXISTING METHODS**

**2.2.1Stock Market Prediction Using Machine Learning**

The research work done by V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India. In the finance world stock trading is one of the most important activities. Stock market prediction is an act of trying to determine the future value of a stock other financial instrument traded on a financial exchange. This paper explains the prediction of a stock using Machine Learning. The technical and fundamental or the time series analysis is used by the most of the stockbrokers while making the stock predictions. The programming language is used to predict the stock market using machine learning is Python. In this paper we propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for the large and small capitalization and in the three different markets, employing prices with both daily and up-to-the-minute frequencies.

**2.2.2Forecasting the Stock Market Index Using Artificial Intelligence Techniques**

The research work done by Lufuno Ronald Marwala A dissertation submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Engineering. The weak form of Efficient Market hypothesis (EMH) states that it is impossible to forecast the future price of an asset based on the information contained in the historical prices of an asset. This means that the market behaves as a random walk and as a result makes forecasting impossible. Furthermore, financial forecasting is a difficult task due to the intrinsic complexity of the financial system. The objective of this work was to use artificial intelligence (AI) techniques to model and predict the future price of a stock market index. Three artificial intelligence techniques, namely, neural networks (NN), support vector machines and neuro-fuzzy systems are implemented in forecasting the future price of a stock market index based on its historical price information. Artificial intelligence techniques have the ability to take into consideration financial system complexities and they are used as financial time series forecasting tools.

Two techniques are used to benchmark the AI techniques, namely, Autoregressive Moving Average (ARMA) which is linear modelling technique and random walk (RW) technique. The experimentation was performed on data obtained from the Johannesburg Stock Exchange. The data used was a series of past closing prices of the All Share Index. The results showed that the three techniques have the ability to predict the future price of the Index with an acceptable accuracy. All three artificial intelligence techniques outperformed the linear model. However, the random walk method out performed all the other techniques. These techniques show an ability to predict the future price however, because of the transaction costs of trading in the market, it is not possible to show that the three techniques can disprove the weak form of market efficiency. The results show that the ranking of performances support vector machines, neuro-fuzzy systems, multilayer perceptron neural networks is dependent on the accuracy measure used.

**2.2.3Indian stock market prediction using artificial neural networks on tick data**

The research work done by Dharmaraja Selvamuthu, Vineet Kumar and Abhishek Mishra Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India. A stock market is a platform for trading of a company’s stocks and derivatives at an agreed price. Supply and demand of shares drive the stock market. In any country stock market is one of the most emerging sectors. Nowadays, many people are indirectly or directly related to this sector. Therefore, it becomes essential to know about market trends. Thus, with the development of the stock

market, people are interested in forecasting stock price. But, due to dynamic nature and liable to quick changes in stock price, prediction of the stock price becomes a challenging task. Stock m Prior work has proposed effective methods to learn event representations that can capture syntactic and semantic information over text corpus, demonstrating their effectiveness for downstream tasks such as script event prediction. On the other hand, events extracted from raw texts lacks of common-sense knowledge, such as the intents and emotions of the event participants, which are useful for distinguishing event pairs when there are only subtle differences in their surface realizations. To address this issue, this paper proposes to leverage external common-sense knowledge about the intent and sentiment of the event.

Experiments on three event-related tasks, i.e., event similarity, script event prediction and stock market prediction, show that our model obtains much better event embeddings for the tasks, achieving 78% improvements on hard similarity task, yielding more precise inferences on subsequent events under given contexts, and better accuracies in predicting the volatilities of the stock market1. Markets are mostly a non- parametric, non-linear, noisy and deterministic chaotic system (Ahangar et al. 2010). As the technology is increasing, stock traders are moving towards to use Intelligent Trading Systems rather than fundamental analysis for predicting prices of stocks, which helps them to take immediate investment decisions. One of the main aims of a trader is to predict the stock price such that he can sell it before its value decline, or buy the stock before the price rises. The efficient market hypothesis states that it is not possible to predict stock prices and that stock behaves in the random walk. It seems to be very difficult to replace the professionalism of an experienced trader for predicting the stock price. But because of the availability of a remarkable amount of data and technological advancements we can now formulate an appropriate algorithm for prediction whose results can increase the profits for traders or investment firms. Thus, the accuracy of an algorithm is directly proportional to gains made by using the algorithm.

**2.2.4The Stock Market and Investment**

The research work done by Manh Ha Duong Boriss Siliverstovs. Investigating the relation between equity prices and aggregate investment in major European countries including France, Germany, Italy, the Netherlands and the United Kingdom. Increasing integration of European financial markets is likely to result in even stronger correlation between equity prices in different European countries. This process can also lead to convergence in economic development across European countries if developments in stock markets influence real economic components, such as investment and consumption. Indeed, our vector autoregressive models suggest that the positive correlation between changes equity prices and investment is, in general, significant. Hence, monetary authorities should monitor reactions of share prices to monetary policy and their effects on the business cycle.

**2.2.5Automated Stock Price Prediction Using Machine Learning**

The research work done by Mariam Moukalled Wassim El-Hajj Mohamad Jaber Computer Science Department American University of Beirut. Traditionally and in order to predict market movement, investors used to analyse the stock prices and stock indicators in addition to the news related to these stocks. Hence, the importance of news on the stock price movement. Most of the previous work in this industry focused on either classifying the released market news as (positive, negative, neutral) and demonstrating their effect on the stock price or focused on the historical price movement and predicted their future movement. In this work, we propose an automated trading system that integrates mathematical functions, machine learning, and other external factors such as news’ sentiments for the purpose of achieving better stock prediction accuracy and issuing profitable trades. Particularly, we aim to determine the price or the trend of a certain stock for the coming end-of-day considering the first several trading hours of the day. To achieve this goal, we trained traditional machine learning algorithms and created/trained multiple deep learning models taking into consideration the importance of the relevant news. Various experiments were conducted, the highest accuracy (82.91%) of which was achieved using SVM for Apple Inc. (AAPL) stock.

**2.2.6Stock Price Correlation Coefficient Prediction with ARIMA- LSTM Hybrid Model**

The research work done by Hyeong Kyu Choi, B.A Student Dept. of Business Administration Korea University Seoul, Korea. Predicting the price correlation of two assets for future time periods is important in portfolio optimization. We apply LSTM recurrent neural networks (RNN) in predicting the stock price correlation coefficient of two individual stocks. RNN’s are competent in understanding temporal dependencies. The use of LSTM cells further enhances its long-term predictive properties. To encompass both linearity and nonlinearity in the model, we adopt the ARIMA model as well. The ARIMA model filters linear tendencies in the data and passes on the residual value to the LSTM model. The ARIMA-LSTM hybrid model is tested against other traditional predictive financial models such as the full historical model, constant correlation model, single-index model and the multi-group model. In our empirical study, the predictive ability of the ARIMA-LSTM model turned out superior to all other financial models by a significant scale. Our work implies that it is worth considering the ARIMALSTM model to forecast correlation coefficient for portfolio optimization.

**2.2.7Event Representation Learning Enhanced with External Common-sense Knowledge**

The research work done by Xiao Ding, Kuo Liao, Ting Liu, Zhongyang Li, Junwen Duan Research Center for Social Computing and Information Retrieval Harbin Institute of Technology, China. Prior work has proposed effective methods to learn event representations that can capture syntactic and semantic information over text corpus, demonstrating their effectiveness for downstream tasks such as script event prediction. On the other hand, events extracted from raw texts lacks of common-sense knowledge, such as the intents and emotions of the event participants, which are useful for distinguishing event pairs when there are only subtle differences in their surface realizations. To address this issue, this paper proposes to leverage external common-sense knowledge about the intent and sentiment of the event. Experiments on three event-related tasks, i.e., event similarity, script event prediction and stock market prediction, show that our model obtains much better event embeddings for the tasks, achieving 78% improvements on hard similarity task, yielding more precise inferences on subsequent events under given contexts, and better accuracies in predicting the volatilities of the stock market.

**2.2.8Forecasting directional movements of stock prices for intraday trading using LSTM and random forests**

The research work done by Pushpendu Ghosh, Ariel Neufeld, Jajati Keshari SahooDepartment of Computer Science & Information Systems, BITS Pilani K.K.Birla Goa campus, India bDivision of Mathematical Sciences, Nanyang Technological University, Singapore cDepartment of Mathematics, BITS Pilani K.K.Birla Goa campus, India. We employ both random forests and LSTM networks (more precisely CuDNNLSTM) as training methodologies to analyse their effectiveness in forecasting out- of-sample directional movements of constituent stocks of the S&P 500 from January 1993 till December 2018 for intraday trading. We introduce a multi-feature setting consisting not only of the returns with respect to the closing prices, but also with respect to the opening prices and intraday returns. As trading strategy, we use Krauss et al. (2017) and Fischer & Krauss (2018) as benchmark and, on each trading day, buy the 10 stocks with the highest probability and sell short the 10 stocks with the lowest probability to outperform the market in terms of intraday returns – all with equal monetary weight. Our empirical results show that the multi-feature setting provides a daily return, prior to transaction costs, of 0.64% using LSTM networks, and 0.54% using random forests. Hence, we outperform the single- feature setting in Fischer & Krauss (2018) and Krauss et al. (2017) consisting only of the daily returns with respect to the closing prices, having corresponding daily returns of 0 .41% and of 0 .39% with respect to LSTM and random forests, respectively. 1 Keywords: Random forest, LSTM, Forecasting, Statistical Arbitrage, Machine learning, Intraday trading.

**2.2.9A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance**

The research work done by Xiao-Yang Liu1 Hongyang Yang,Qian Chen4,Runjia ZhangLiuqing Yang Bowen Xiao Christina Dan Wang Electrical Engineering, 2Department of Statistics, 3Computer Science, Columbia University, 3AI4Finance LLC., USA, Ion Media Networks, USA, Department of Computing, Imperial College, 6New York University (Shanghai). As deep reinforcement learning (DRL) has been recognized as an effective approach in quantitative finance, getting hands-on experiences is attractive to beginners. However, to train a practical DRL trading agent that decides where to trade, at what price, and what quantity involves error-prone and arduous development and debugging. In this paper, we introduce a DRL library FinRL that facilitates beginners to expose themselves to quantitative finance and to develop their own stock trading strategies. Along with easily-reproducible tutorials, FinRL library allows users to streamline their own developments and to compare with existing schemes easily.

Within FinRL, virtual environments are configured with stock market datasets, trading agents are trained with neural networks, and extensive back testing is analysed via trading performance. Moreover, it incorporates important trading constraints such as transaction cost, market liquidity and the investor’s degree of risk-aversion. FinRL is featured with completeness, hands-on tutorial and reproducibility that favors beginners:

(i) at multiple levels of time granularity, FinRL simulates trading environments across various stock markets, including NASDAQ-100, DJIA, S&P 500, HSI, SSE 50, and CSI 300; (ii) organized in a layered architecture with modular structure, FinRL provides fine- tuned state-of-the-art DRL algorithms (DQN, DDPG, PPO, SAC, A2C, TD3, etc.), commonly used reward functions and standard evaluation baselines to alleviate the debugging workloads and promote the reproducibility, and (iii) being highly extendable, FinRL reserves a complete set of user-import interfaces. Furthermore, we incorporated three application demonstrations, namely single stock trading, multiple stock trading, and portfolio allocation. The FinRL library will be available on GitHub at link https://github.com/AI4Finance-LLC/FinRL-Library.

**2.2.10An innovative neural network approach for stock market prediction**

The research work done by Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin. To develop an innovative neural network approach to achieve better stock market predictions. Data were obtained from the live stock market for real-time and off-line analysis and results of visualizations and analytics to demonstrate Internet of Multimedia of Things for stock analysis. To study the influence of market characteristics on stock prices, traditional neural network algorithms may incorrectly predict the stock market, since the initial weight of the random selection problem can be easily prone to incorrect predictions.

Based on the development of word vector in deep learning, we demonstrate the concept of “stock vector.” The input is no longer a single index or single stock index, but multi-stock high-dimensional historical data. We propose the deep long short-term memory neural network (LSTM) with embedded layer and the long short-term memory neural network with automatic encoder to predict the stock market. In these two models, we use the embedded layer and the automatic encoder, respectively, to vectorize the data, in a bid to forecast the stock via long short-term memory neural network. The experimental results show that the deep LSTM with embedded layer is better. Specifically, the accuracy of two models is 57.2 and 56.9%, respectively, for the Shanghai A-shares composite index. Furthermore, they are 52.4 and 52.5%, respectively, for individual stocks. We demonstrate research contributions in IMMT for neural network-based financial analysis. 2.2.11 An Intelligent Technique for Stock Market Prediction

**2.2.11An Intelligent Technique for Stock Market Prediction**

The research work done by M. Mekayel Anik · M. Shamsul Arefin

(B) Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh. A stock market is a loose network of economic transactions between buyers and sellers based on stocks also known as shares. In stock markets, stocks represent the ownership claims on businesses. These may include securities listed on a stock exchange as well as those only traded privately. A stock exchange is a place where brokers can buy and/or sell stocks, bonds, and other securities. Stock market is a very vulnerable place for investment due to its volatile nature. In the near past, we faced huge financial problems due to huge drop in price of shares in stock markets worldwide. This phenomenon brought a heavy toll on the international as well as on our national financial structure. Many people lost their last savings of money on the stock market. In 2010–2011 financial year, Bangladeshi stock market faced massive collapse [1]. This phenomenon can be brought under control especially by strict monitoring and instance stock market analysis. If we can analyse stock market correctly in time, it can become a field of large profit and may become comparatively less vulnerable for the investors.

Stock market is all about prediction and rapid decision making about investment, which cannot be done without thorough analysis of the market. If we can predict the stock market by analysing historical data properly, we can avoid the consequences of serious market collapse and to be able to take necessary steps to make market immune to such situations.

# Chapter 3

**System Analysis**

**3.1 Proposed System**

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 machine-learning algorithm is to perform the best predicting result of the stock future price. LSTM is capable to catching the modifications in the behaviour of the stock price for the indicated period in this proposed system. a machine learning-based normalization for stock price prediction. The dataset utilized for analysis was selected from Yahoo Finance. It consists of approximately 9 lakh records of the required Stock price and other relevant data. The data reflected the stock price at some time intervals for every day of the year. It contains various data like date, symbol, open price, close price, low price, high price and volume. Here, the data for only one company was considered. All the data was available from yahoo finance website which was first read and transformed into a data frame using the Pandas library in Python. The normalization of the data was performed through the sklearn library in Python and the data were divided into training and testing sets. The experiment set was kept as 20% of the available dataset. This paper focuses on two architecture Regression-based Model and LSTM. The Regression-based Model is employed for predicting unbroken values through some given autonomous values Regression uses a given linear function for predicting continuous values of the most important amongst them and made the predictions using these. LSTM architecture is able to identify the changes in trends which show evident from the result. LSTM is identified as the best model for the proposed methodology. This shows that the proposed system is capable of identifying some interrelation within the data. In the stock market, there may not always follow the same cycle or may not always be in a regular pattern for the changes that are occurred. The period of the existence will differ and the existence of the trend is based on the companies and the sectors. For investors, this type of analysis of trends and cycles will obtain more profit. We must use networks like LSTM as they rely on the current information to analyse various information.

3.1.1 proposed work

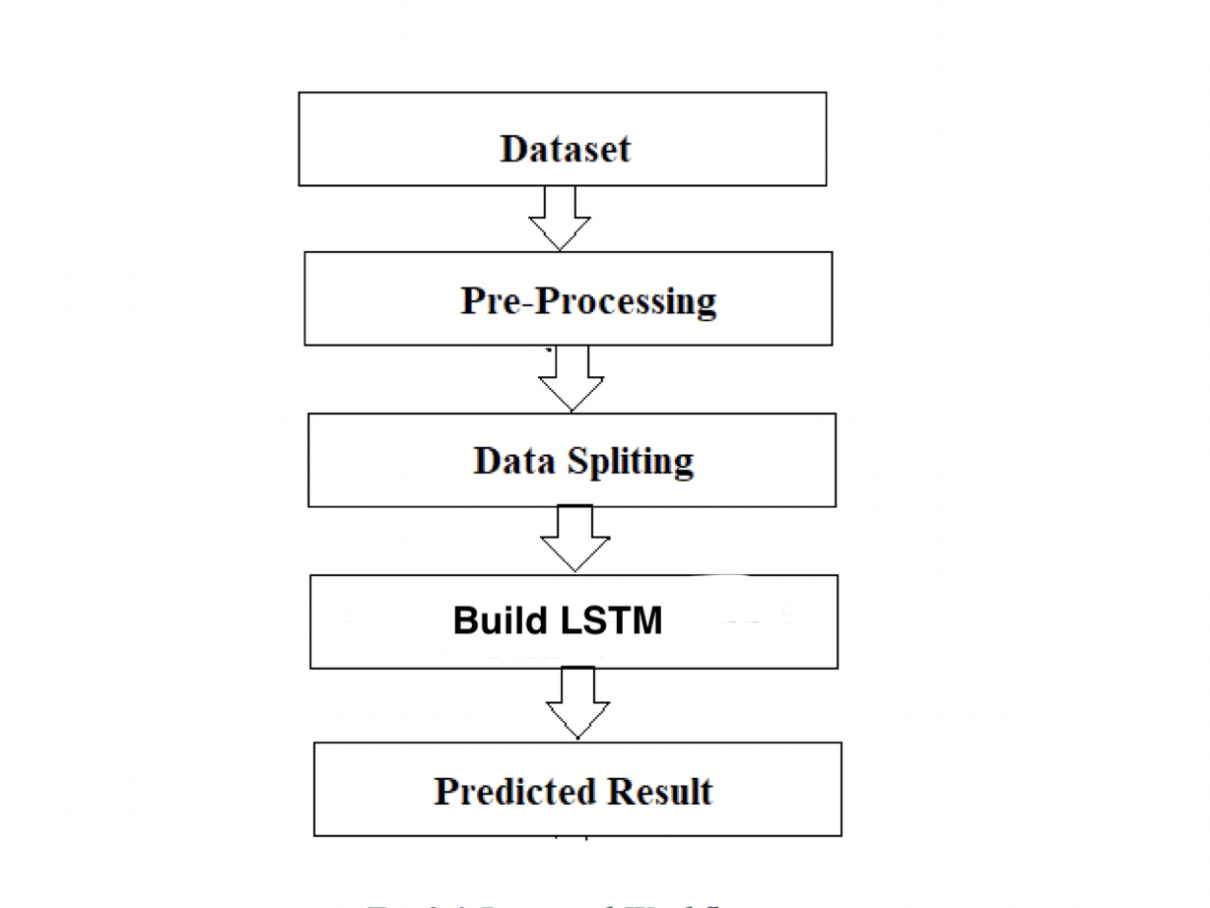


Fig 1.1 Work flow

# The system presented here composes of five modules:-

1. Input as Dataset

# 2. Pre processing

# 3. Data splitting

# 4. Build & Model train Lstm

# 5. Output as Predicted Result

Attribute such as: price of open, high, low, close, adjusted close price taken from huge dataset are fed as input to the models for training to pre-process the data techniques like normalization & one hot encoding in applied on dataset. After this data is divided in two sets namely training & testing which are ratio of 70:30 respectively. Then, this set are used to train a model using LSTM approach.Finally, all these modules are evaluated using Root mean square error.

**3.2 SYSTEM ARCHITECTURE**

**3.2.1 Preprocessing of data**

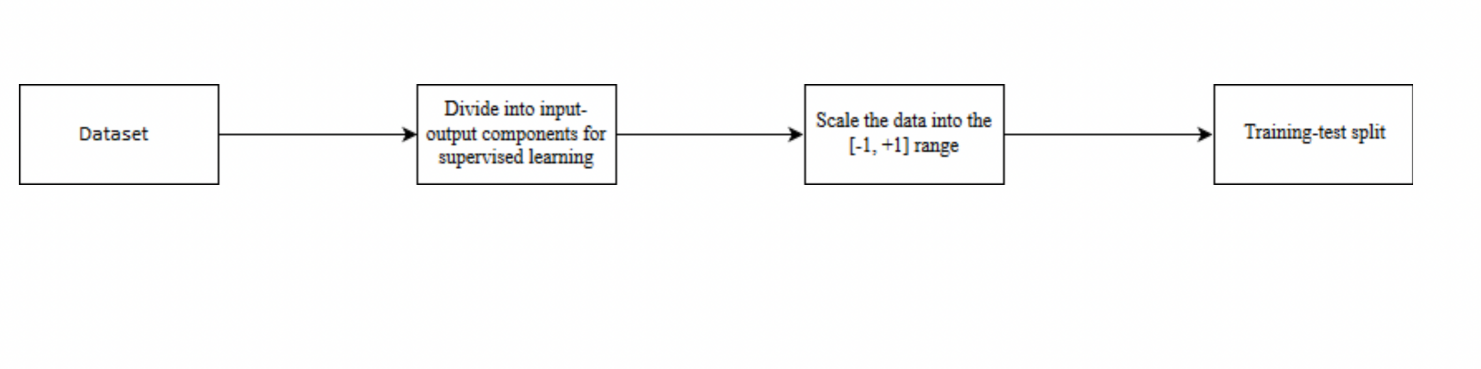
****

Fig 1.2 Data Preprocessing

**3.2.2 overall architecture**

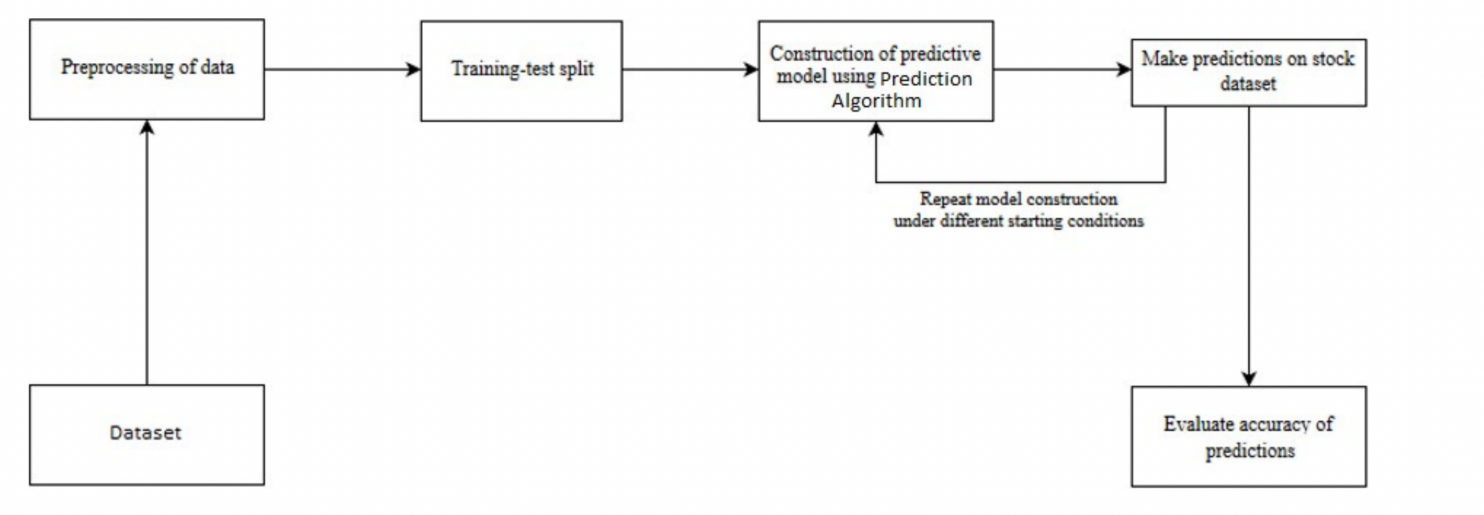


Fig 1.3 Overall Architecture

Chapter 4

# System Design And Implementation

**4.1 Introduction:**

**Long short-term memory network:**

# Long short-term memory network (LSTM) is a particular form of recurrent neural network (RNN).

**4.1.1 Working of LSTM:**

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.

•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%.

**4.2 Hardware Requirements:**

# RAM: 4 GB

* Storage: 500 GB
* CPU: 2 GHz or faster
* Architecture: 32-bit or 64-bit

**4.3 Software Requirements:**

* Python 3.5 in Google Colab is used for data pre-processing, model training and prediction.
* Operating System: windows 7 and above or Linux based OS or MAC OS

**4.4 Dataset Detail:**

The dataset consists of the stock historical data from the National stock exchange (NSE) and captures the daily information of each stock from the National Stock Exchange. It collects different sectors of stock data, including Banking, Pharma, Petroleum, Software and Textiles and it including the opening price, the highest price, the lowest price, the closing price, the adjusted closing price and the volume of stock.

# 4.5 Tool & Technologies

# 4.5.1 PYTHON

The language of select for this project was Python. This was a straightforward call for many reason.

Python as a language has a vast community behind it. Any problems which may be faced is simply resolved with visit to Stack Overflow. Python is the foremost standard language on the positioning that makes it is very straight answer to any question.Python is an abundance of powerful tools ready for scientific computing Packages. The packages like NumPy, Pandas and SciPy area unit freely available and well documented. These Packages will intensely scale back, and variation the code necessary to write a given program. This makes repetition fast. Python is a language as forgiving and permits for the program that appear as if pseudo code. This can be helpful once pseudo code give in tutorial papers should be required and verified. Using python this step is sometimes fairly trivial. However, Python is not without its errors. The python is dynamically written language and packages are area unit infamous for Duck writing. This may be frustrating once a package technique returns one thing that, for instance, looks like an array instead of being an actual array. Plus the standard Python documentation did not clearly state the return type of a method, this can’t lead without a lot of trials and error testing otherwise happen in a powerfully written language. This is a problem that produces learning to use a replacement Python package or library more difficult than it otherwise may be.

**4.5.2 NUMPY**

Numpy is python package which provide scientific and higher level mathematical abstractions wrapped in python. It is the core library for scientific computing, that contains a provide tools for integrating C, strong n-dimensional array object, C++ etc. It is also useful in random number capability, linear algebra etc.Numpy’s array type augments the Python language with an efficient data structure used for numerical work, e.g., manipulating matrices. Numpy additionally provides basic numerical routines, like tools for locating Eigenvectors

**4.5.3 SCIKIT LEARN**

Scikit-learn could be a free machine learning library for Python. It features numerous classification, clustering and regression algorithms vector machine, and it furthermore supports Python scientific and numerical libraries like SciPy and NumPy. In Python Scikit-learn is specifically written, with the core algorithms written in Cython to get the performance. Support vector machines are enforced by a Cython wrapper around LIBSVM .i.e., linear support vector machines and logistic regression by a similar wrapper around LIBLINEAR.

**4.5.4 TENSORFLOW**

like random forests, k-neighbours, support In the TensorFlow [22]has an open source software library for numerical computation using data flow graphs. Inside the graph nodes represent mathematical formulae, the edges of graph represent the multidimensional knowledge arrays (tensors) communicated between them. The versatile architecture permits to deploy the computation to at least one or many GPUs or CPUs in a desktop, mobile device, servers with a single API. TensorFlow was firstly developing by engineers and researchers acting on the Google Brain Team at intervals Google's Machine Intelligence analysis organization for the needs of conducting deep neural networks research and machine learning, but, the system is generally enough to be appropriate in a wide range of alternate domains as well. Google Brain's second-generation system is TensorFlow. Whereas the reference implementation runs on single devices, TensorFlow can run on multiple GPUs and CPUs. TensorFlow is offered on Windows, macOS, 64-bit Linux and mobile computing platformstogether with iOS and Android

# 4.5.5 KERAS

# Keras is a high-level neural networks API, it is written in Python and also capable of

# running on top of the Theano, CNTK, or. TensorFlow. It was developed with attention on enabling quick experimentation. having the ability to travel from plan to result with the smallest amount doable delay is key to doing great research.Keras permits for straightforward and quick prototyping (through user-friendliness, modularity, and extensibility). Supports each recurrent networks and convolutional networks, also as combinations of the 2. Runs seamlessly on GPU and CPU. The library contains numerous implementations of generally used neural network building blocks like optimizers, activation functions, layers, objectives and a number of tools to create operating with text and image data easier. The code is hosted on GitHub, and community support forums embody the GitHub issues page, a Gitter channel

# and a Slack channel.

# 4.6 implementation

# 4.6.1 Read Dataset:

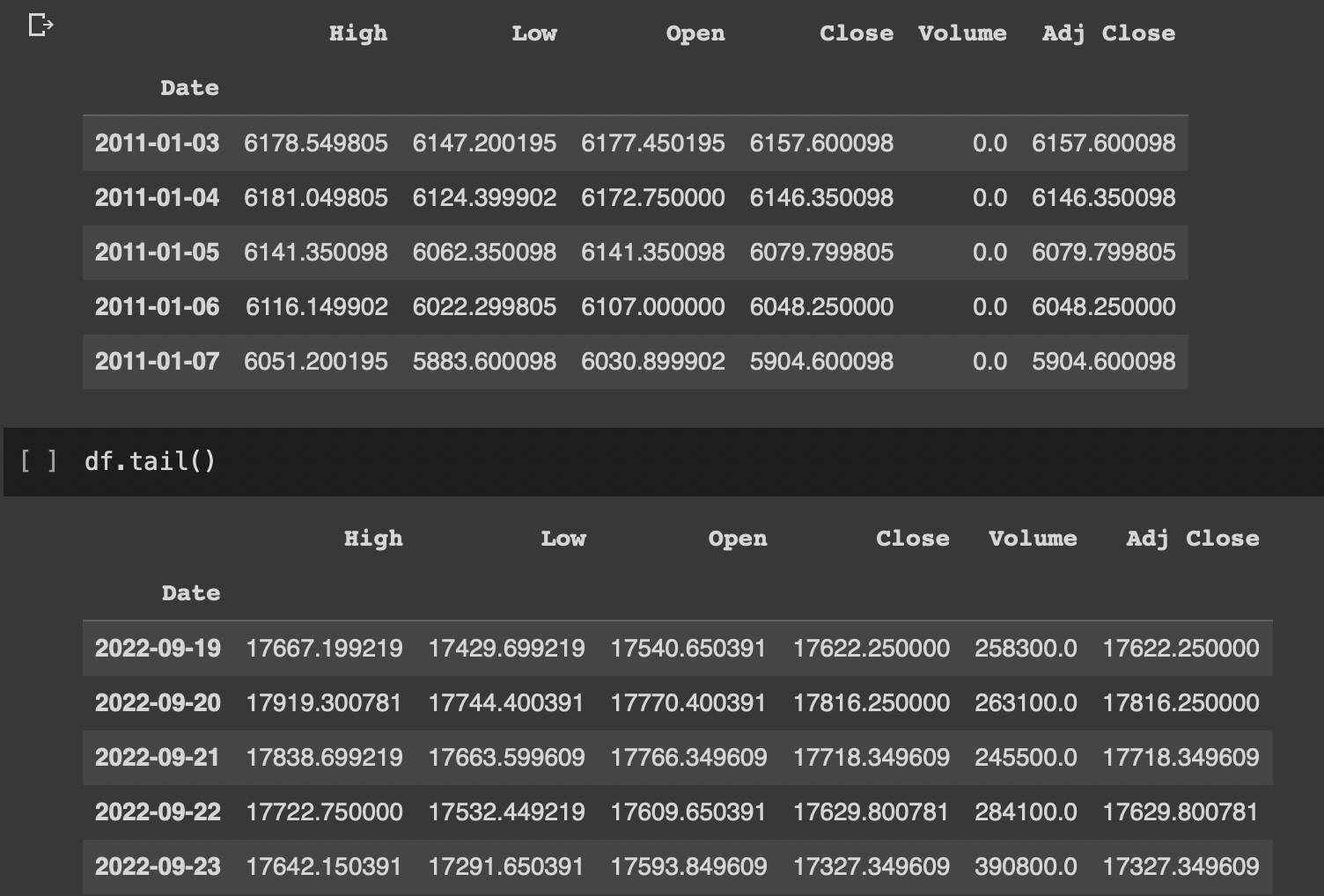


Fig 2.1 Dataset

# 4.6.2 graph of close history

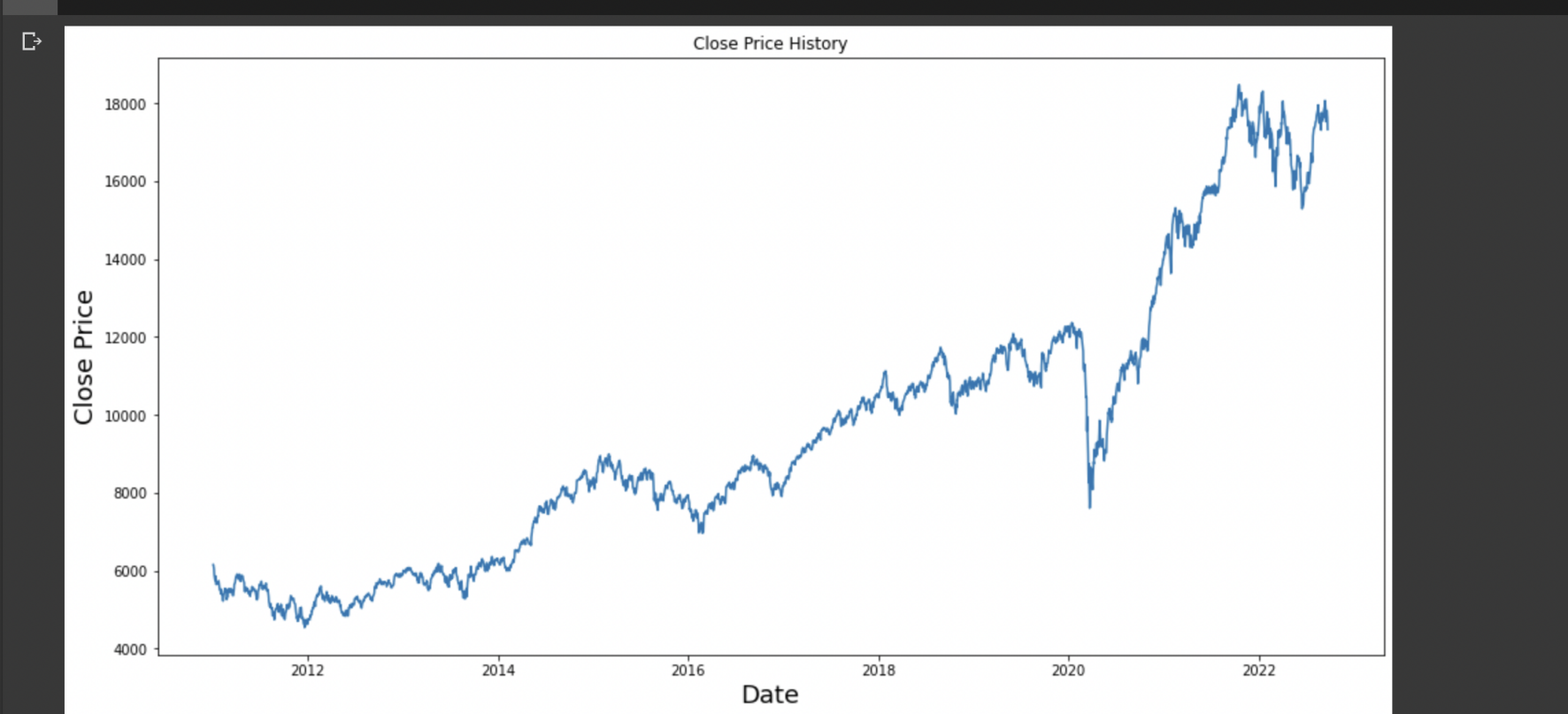


Fig 2.2 Close History Graph

# 4.6.3 model fitting summary

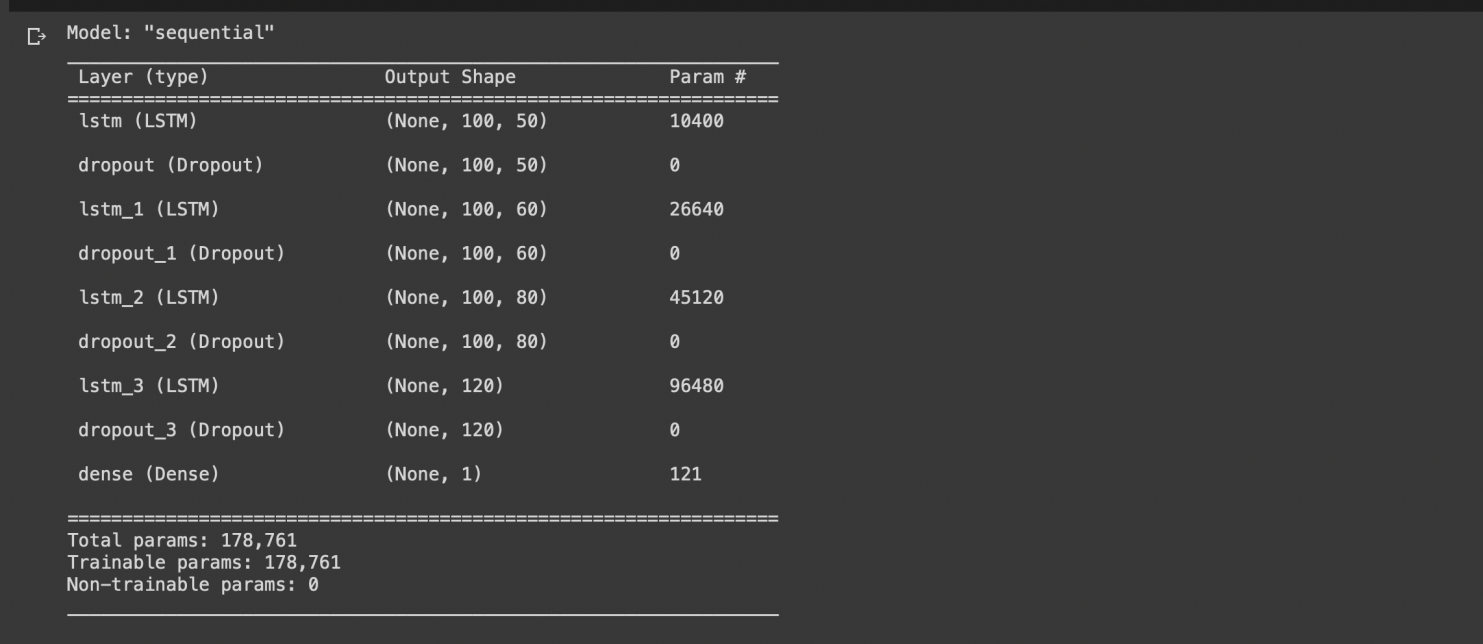


Fig 2.3 Model Summary

# 4.6.4 prediction

# Screenshot 2022-09-29 at 15.51.41

Fig 2.3 Prediction

**Chapter 5**

# Performance analysis

**5.1system configuration**

we ran entire software on google collaboratory.

**5.1.1 Hardware Requirements:**

* RAM: 4 GB
* Storage: 500 GB
* CPU: 2 GHz or faster
* Architecture: 32-bit or 64-bit

**5.1.2 Software requirements**

* Python 3.8 in Google Colab is used for data pre-processing, model training and prediction.
* Operating System: windows 7 and above or Linux based OS or MAC OS.

# 5.2 LSTM NIFTY50

|  |  |
| --- | --- |
| MSE | MAE |
| 377304.1486660041 | 475.803996521377 |

**Chapter 6**

Fig 3 LSTM Performance

**FUTURE ENHANCEMENT AND CONCLUSION**

**6.1 limitations of the system:**

* Not able to update the dataset automatically, we have to manually do it.
* Overall UI of the webApp is not upto the mark.

**6.2 Conclusion**

In this project, we are predicting closing stock price of any given

organization, we developed a web application for predicting close stock price using LSTM algorithms for prediction. We have applied datasets belonging to NIFTY50 and achieved above 95% accuracy for this

dataset.

**6.3 Future enhancements:**

* we will compare a machine learning models like LSTM model, the CNN model and also the hybrid approach of LSTM + CNN model
* In the future, for better accuracy model can be trained with more varied and detailed data. Also, other algorithms along with proposed can be used to create a new hybrid model.
* We want to extend this application for predicting cryptocurrency trading.

**References**

1. Stock Price Prediction Using LSTM on Indian Share Market by Achyut Ghosh, Soumik Bose1,

GiridharMaji, Narayan C. Debnath, Soumya Sen

1. S. Selvin, R. Vinayakumar, E. A. Gopalkrishnan, V. K. Menon and K. P. Soman - Stock price predictionusing LSTM, RNN and CNN-sliding window model - 2017.
2. Murtaza Roondiwala, Harshal Patel, Shraddha Varma, “Predicting Stock Prices Using LSTM” in Undergraduate Engineering Students, Department of Information Technology, Mumbai University, 2015.
3. Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, “An innovative neural network approach for stock market prediction”, 2018
4. Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, Lokesh Chouhan Department of Computer Science and Engineering National Institute of Technology, Hamirpur – 177005, INDIA - Stock Market Prediction Using Machine Learning.
5. Pranav Bhat Electronics and Telecommunication Department, Maharashtra Institute of Technology, Pune. Savitribai Phule Pune University - A Machine Learning Model for Stock Market Prediction.
6. Anurag Sinha Department of computer science, Student, Amity University Jharkhand Ranchi, Jharkhand (India), 834001 - Stock Market Prediction Using Machine Learning.
7. V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India

- Stock Market Prediction Using Machine Learning.