

**department of Computer Engineering & Applications**

**INSTITUTE OF ENGINEERING & TECHNOLOGY**

**B.Tech. IV Year CSE**

**Project Report**

**On**

**“Stock Price Prediction using LSTM and News Sentiment Analysis”**

**Under the supervision of** **Submitted by**

**1. Divyam Gupta (26 / 181500219)**

**Dr. Swati Srivastava 2. Rishabh Tiwari (28 / 181500566)**

**3. Rohan Bharadwaj (52 / 181500589)**

**Group No: G-37**



**Declaration**

We hereby declare that the work which is being presented in the Major Project I on **“**Stock Price Prediction using LSTM and News Sentiment Analysis**”,** in partial fulfilment of the requirements for Major Project viva, is an authentic record of our own work carried under the supervision of **Dr. Swati Srivastava, Assistant Professor, GLA University, Mathura.**

Divyam Gupta (181500219)

Sign: \_\_\_\_\_\_\_\_\_\_\_\_\_

Rishabh Tiwari (181500566)

Sign: \_\_\_\_\_\_\_\_\_\_\_\_\_

Rohan Bharadwaj (181500589)

Sign: \_\_\_\_\_\_\_\_\_\_\_\_\_



**Certificate**

This is to certify that the project entitled “Stock Price Prediction using LSTM and News Sentiment Analysis” carried out in Major Project I is the work done by Divyam Gupta, Rishabh Tiwari and Rohan Bharadwaj is submitted in partial fulfilment of the requirements for the award of degree Bachelor of Technology (Computer Science and Engineering).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Supervisor**

(**Name of Supervisor**)

Designation of Supervisor

Dept. of Computer Engg, & App.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Project Co-ordinator Program Co-ordinator**

**(Dr. Mayank Srivastava)** (**Dr Rakesh Kumar Galav**)

Associate Professor Assistant Professor

Dept. of Computer Engg, & App. Dept. of Computer Engg, & App.

Date:



**Abstract**

The stock market is a field of study for researchers because of its sudden changes and difficult to predict nature. The stocks of any company are affected by its performance and external factors like government policies, reaction of the investors. The aim is to predict the price of stocks or the stock movement for a given day using historic stock data and recent news. Two different models, one for historic data and other for news sentiment are trained separately. The stock data after pre-processing is used to extract pattern using long short term memory (LSTM) algorithm which uses feedback connections. The other model is using natural language processing to process the sentiments of news about particular stocks which will affect the stock prices. From the second models, a bag of words is extracted which will help in deciding the direction and impact of news on stock movement. Thus, a final prediction will be made using the above two models.

**Index**

Chapter 1: Introduction………………………………………………………………………6

* 1. Overview and Motivation………………………………………………………..6
  2. Objective…………………………………………………………………………7
  3. Scope……………………………………………………………………………..7

Chapter 2: Literature Survey…………………………………………………………………8

2.1 Introduction……………………………………………………………………………….12  
2.2 Existing methods………………………………………………………………………….12  
 2.2.1 Stock Market Prediction Using Machine Learning…………………………….12  
 2.2.2 Forecasting the Stock Market Index Using Artificial Intelligence  
 Techniques…………………………………………………………………....12  
 2.2.3 Indian stock market prediction using artificial neural network…………….....13  
 2.2.4 The Stock Market and Investment………………………………………….....14  
 2.2.5 Automated Stock Price Prediction Using Machine Learning………………….15  
 2.2.6 Stock Price Correlation Coefficient Prediction with ARIMA-  
 LSTM Hybrid Model………………………………………………………….15  
 2.2.7 Event Representation Learning Enhanced with External Common-  
 sense Knowledge………………………………………………………………16  
 2.2.8 Forecasting directional movements of stock prices for trading  
 using LSTM and random forests………………………………………………16  
 2.2.11An Intelligent Technique for Stock Market Prediction………………………..17

Chapter 3: Proposed Model…………………………………………………………………..19

* 1. Flow Diagram…………………………………………………………………….19
  2. Model-I…………………………………………………………………………...20

3.2.1MinMaxScaler…………………………………………………………..21

3.2.2 Mean Squared Error……………………………………………………22

3.2.3 Adam Optimizer………………………………………………………..22

3.2.4 LSTM…………………………………………………………………...23

* 1. Model-II…………………………………………………………………………..26

Chapter 4 : Experimental Analysis And Results……………………………………………..28

4.1 System Configuration…………………………………………………………………………………….……………. 28

4.1.1 Hardware Requirements………………………………………………………….…………..28

4.1.2 Software Requirements……………………………………………………………..…………28

**4.2** Sample Code…………………………………………………………………………………………………………………29

4.3 Sample Output………………………………………………………………………………………………………..…….32

References…………………………………………………………………………………….33

**Chapter 1**

**Introduction**

* 1. **Overview and Motivation**

A stock market is a public market where you can buy and sell shares for publicly listed companies. The stocks, also known as equities, represent ownership in the company. The stock exchange is the mediator that allows the buying and selling of shares. The stock market has its importance both for personal and professionals like it help companies to raise capital, it helps generate personal wealth, Stock markets serve as an indicator of the state of the economy, it is a widely used source for people to invest money in companies with high growth potential. The stock market plays a remarkable role in our daily lives. It is a significant factor in a country's GDP growth.

Stock market is an investment place for business people and also for the common people. It can increase the invested value or can make it null in a minute. The stock market has its ups and down throughout any period of time which makes it difficult to predict. The movement of any stock is affected by that company itself as an internal factor like performance, etc. but also due to external factors like investor’s behavior, etc. We have proposed a model to understand and predict the stock prices on daily basis. The prediction is based on the historical stock data and recent news which can tell the market trend.

Stock market prediction and analysis are some of the most difficult jobs to complete. There are numerous causes for this, including market volatility and a variety of other dependent and independent variables that influence the value of a certain stock in the market. These variables make it extremely difficult for any stock market expert to anticipate the rise and fall of the market with great precision.

However, with the introduction of Machine Learning and its strong algorithms, the most recent market research and Stock Market Prediction advancements have begun to include such approaches in analyzing stock market data.

The motivation behind this project is by knowing the trend of the market and including other factors that can affect it, one is able to predict how it will behave in coming future. The successful prediction of a stock's future price could yield significant profit. It’s a classic problem and still in the research.

* 1. **Objective**

The main objective is to predict the closing price of stocks on day basis using a GUI for input of required data and to show the final result. Trying to determine the future value of a company stock or other financial instruments traded on an exchange. The successful prediction of a stock's future price could yield significant profit.

* 1. **Scope**
* Analysis and prediction of stocks will be useful for new investors to invest in stock market based on the various factors considered by the software.
* Analyzing stock market with the news.

**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.

**2.2 EXISTING METHODS**

**2.2.1 Stock 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 capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies.

**2.2.2 Forecasting 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.

**2.2.3 Indian 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.

**2.2.4 The 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.

**2.2.5 Automated 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.

**2.2.6 Stock 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.

**2.2.7 Event 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.

**2.2.8 Forecasting 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.

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 financialanalysis.

**2.2.9 An 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].

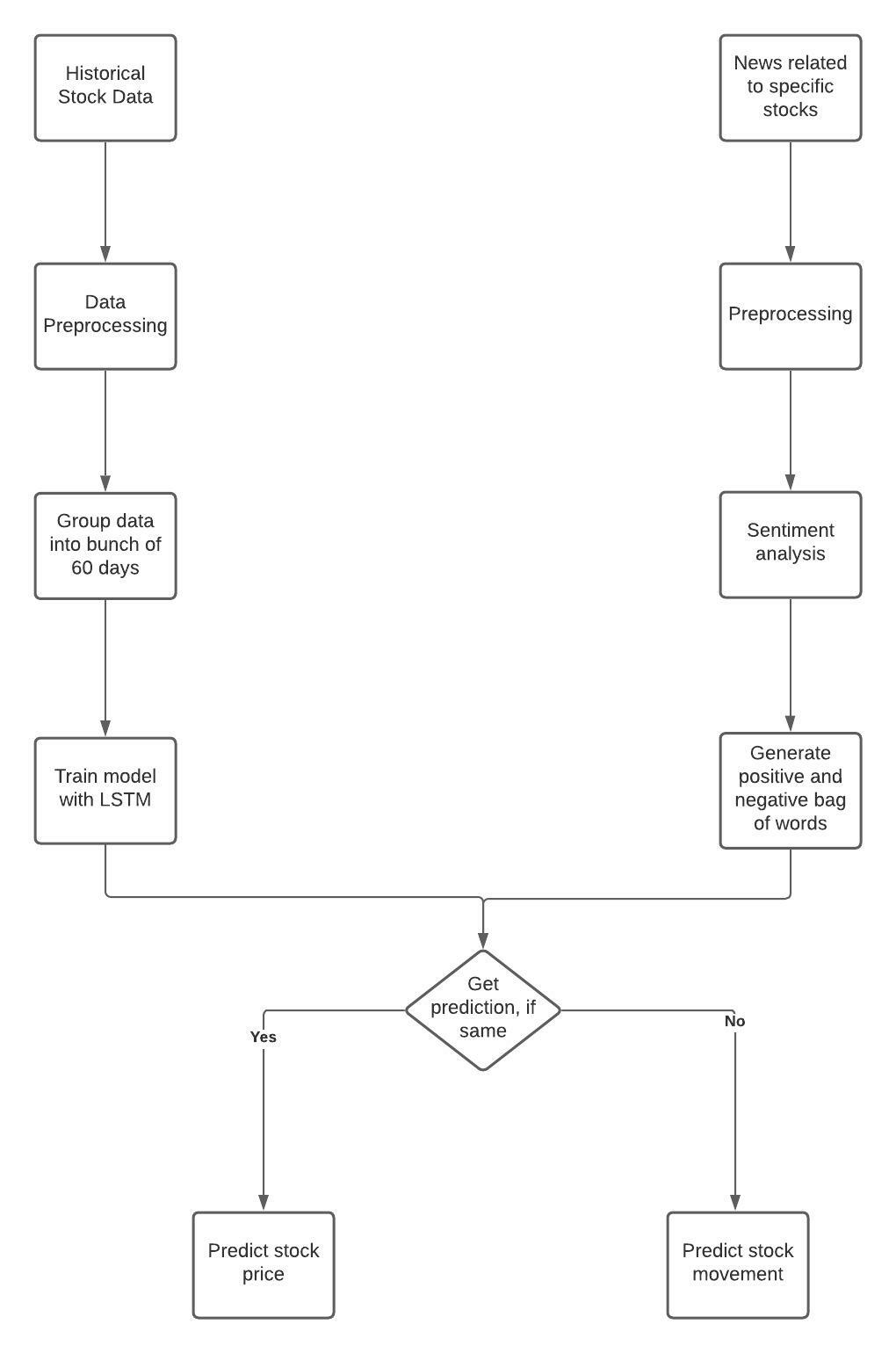
This phenomenon can be brought under control especially by strict monitoring and instance stock market analysis. If we can analyse stock market correctly in time, it can become a field of large profit and may become comparatively less vulnerable for the investors. Stock market is all about prediction 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**

**Proposed Model**

**3.1 Flow Diagram**

****

The model we proposed contains two models, one for historical price data and another for news sentiment analysis.

**3.2 Model-I**

The first model will deal with the actual stock prices of a company. Dataset used here is being Yahoo finances with duration of last 5 years. An example of dataset:

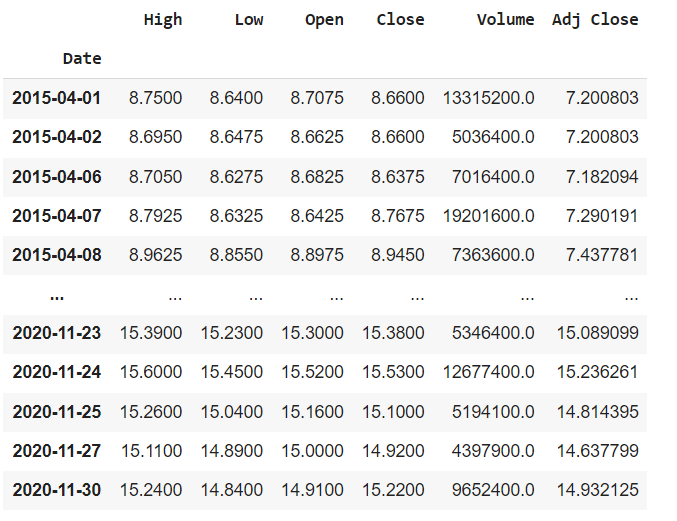


Fig 1: Example of Dataframe

For initial, the closing prices of a company are mapped with dates to build a predictive model.

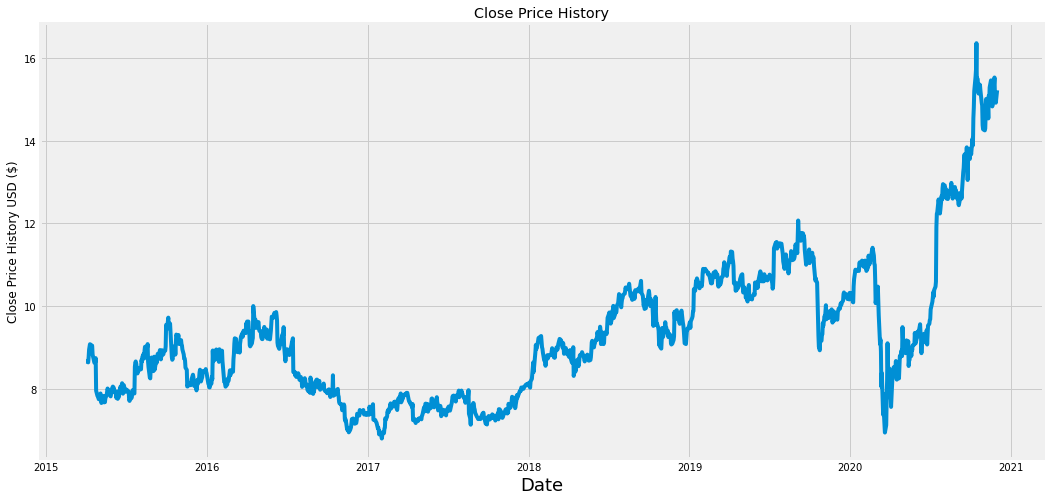


Fig 2: A map between closing price and date of stock

The 20% data of last days is taken for testing and the rest 80% will be used for training. The training dataset is first preprocessed with MinMaxScaler to scale the whole data between 0 and 1 for uniformity and better data capturing.

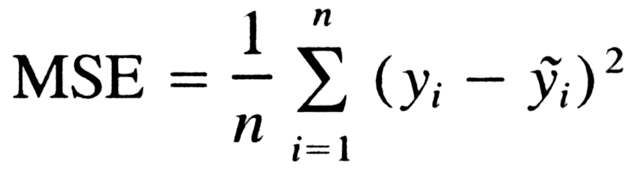
**3.2.1 MinMaxScaler**

Itremoves the mean and scales each feature/variable to unit variance. This operation is performed feature-wise in an independent way. Normally it is used when lower and upper boundaries for scaling are known.

The model detects a pattern from a 60-day dataset and gives prediction for the 61st day. So, the whole dataset is divided into group of 60-days like from 1st to 60th row, 2nd to 61st row and so on. Now these groups of data will become input for a LSTM model consisting of four different layers of neural networks. The error while training the model will be calculated through mean-squared error using the Adam optimizer.

**3.2.2 Mean Squared Error**

The mean squared error (MSE) or mean squared deviation (MSD) of an estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors, i.e. the average squared difference between the estimated values and the actual values. MSE is a risk function, corresponding to the expected value of the squared error loss. The MSE is a measure of the quality of an estimator. As it is derived from the square of Euclidean Distance, it is always a positive value that decreases as the error approaches zero. The MSE is the second moment (about the origin) of the error, and thus incorporates both the variance of the estimator (how widely spread the estimates are from one data sample to another) and its bias (how far off the average estimated value is from the true value). It is calculated by the formula:



**3.2.3 Adam Optimizer**

Adaptive Moment Estimation is an algorithm for optimization technique for gradient descent. It involves a combination of two gradient descent methodologies:

* One is momentum which is used to accelerate the gradient descent algorithm by taking into consideration the ‘exponentially weighted average’ of the gradients. Using averages makes the algorithm converge towards the minima in a faster pace.
* And the other is root mean square propagation (RMSP), is an adaptive learning algorithm that tries to improve AdaGrad. Instead of taking the cumulative sum of squared gradients like in AdaGrad, it takes the ‘exponential moving average’.

Adam Optimizer inherits the strengths or the positive attributes of the above two methods and builds upon them to give a more optimized gradient descent.

Here, we control the rate of gradient descent in such a way that there is minimum oscillation when it reaches the global minimum while taking big enough steps (step-size) so as to pass the local minima hurdles along the way. Hence, combining the features of the above methods to reach the global minimum efficiently.

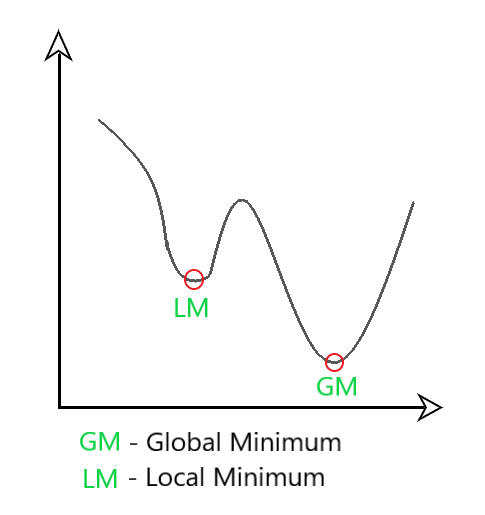


Fig 3: Local and Global Minimum

**3.2.4 Long Short Term Memory (LSTM)**

It is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike normal feed forward neural network, LSTM has feedback connections. It can process not only single data points (such as images), but also entire sequences of data (such as speech and video). A common LSTM unit or neuron is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs. A single LSTM cell looks like:

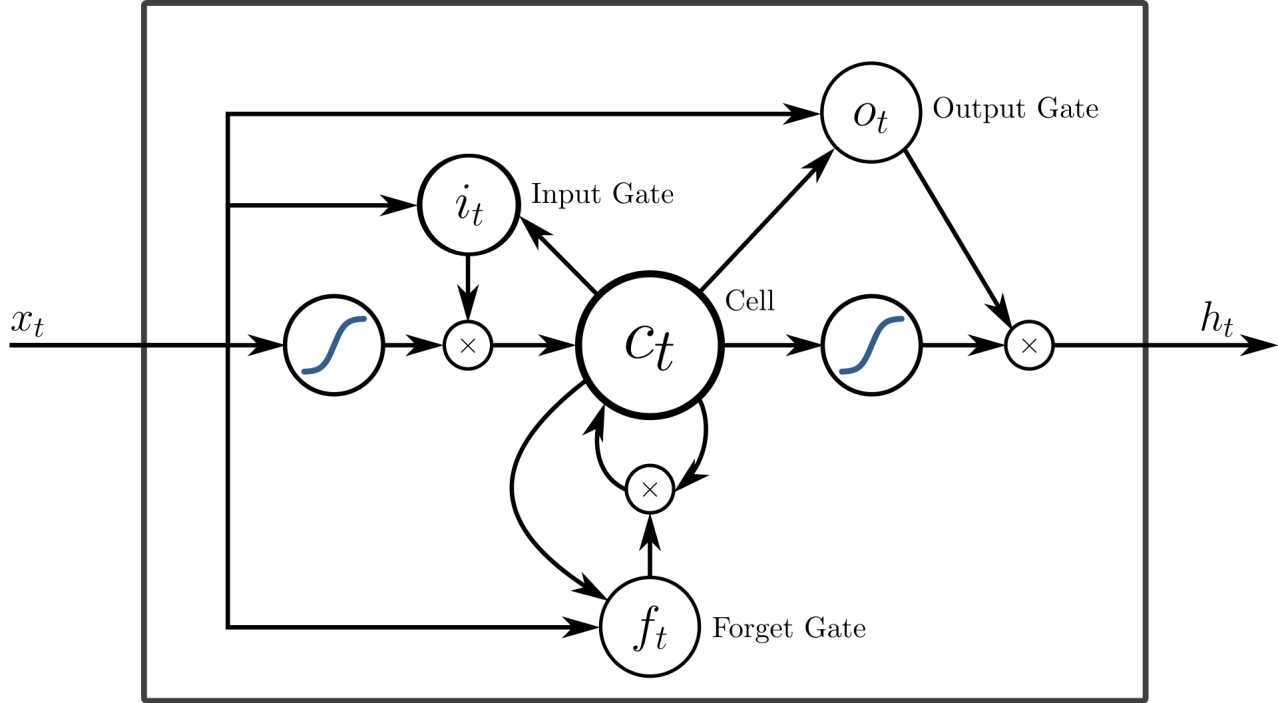


Fig 4: A Peephole Unit in LSTM

Each of the gates can be thought as a "standard" neuron in a feed-forward (or multi-layer) neural network: that is, they compute an activation (using an activation function) of a weighted sum. it, ot and ft represent the activations of respectively the input, output and forget gates, at time step *t*.

The 3 exit arrows from the memory cell *c* to the 3 gates *i, o* and *f* represent the peephole connections. These peephole connections actually denote the contributions of the activation of the memory cell *c* at time step *t-1*, i.e. the contribution of ct-1 (and not ct, as the picture may suggest). In other words, the gates *i*, *o*, and *f* calculate their activations at time step *t* (i.e., respectively, it, ot and ft) also considering the activation of the memory cell *c* at time step *t-1*, i.e. *ct-1*.

The single left-to-right arrow exiting the memory cell is not a peephole connection and denotes *ct*.

The little circles containing a *x* symbol represent an element-wise multiplication between its inputs. The big circles containing an S-like curve represent the application of a differentiable function (like the sigmoid function) to a weighted sum.

LSTMs work in a three-step process.

* The first step in LSTM is to decide which information to be omitted from the cell in that particular time step. It is decided with the help of a sigmoid function. It looks at the previous state (ht-1) and the current input xt and computes the function.
* There are two functions in the second layer. The first is the sigmoid function, and the second is the tanh function. The sigmoid function decides which values to let through (0 or 1). The tanh function gives the weightage to the values passed, deciding their level of importance from -1 to 1.
* The third step is to decide what will be the final output. First, you need to run a sigmoid layer which determines what parts of the cell state make it to the output. Then, you must put the cell state through the tanh function to push the values between -1 and 1 and multiply it by the output of the sigmoid gate.

Now, using the LSTM, the model is trained on dataset with an RMSE of 0.147. A summary of model used:

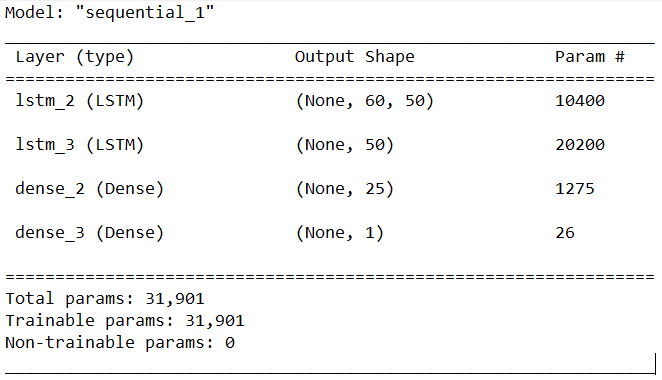


Fig 5: LSTM Model Summary

The model has used 50 neurons for each first and second LSTM layer, then 25 neurons in first dense and one neuron as last layer which will give final result.

On testing the model with the testing dataset:

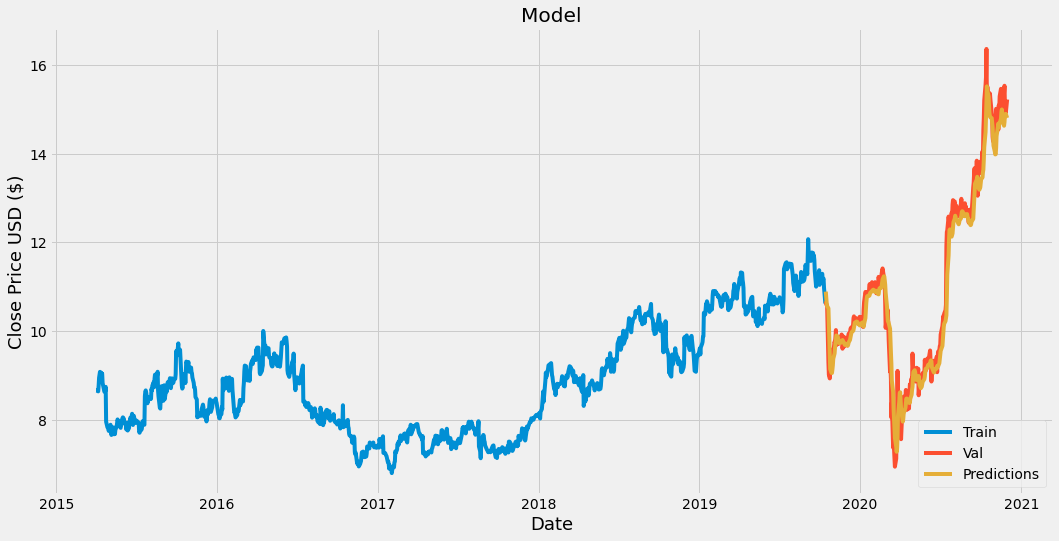


Fig 6: A map between predicted and actual values

**3.3 Model-II**

The second model uses the python news API client for fetching news dataset. To perform the sentiment analysis on the news, firstly it should be pre-processed which includes removal of unnecessary things like commas, special characters, etc. Then with the polarity score which lies between -1.0 to 1.0, we can find polarity of statements. These news statements with polarity scores will be used to train and observe the impact of news on the stock movement.

**Polarity Score:** Sentiment polarity for an element defines the orientation of the expressed sentiment, i.e., it determines if the text expresses the positive, negative or neutral sentiment of the user about the entity in consideration. It lies between -1.0 and 1.0 and a float value.

An example of news data:

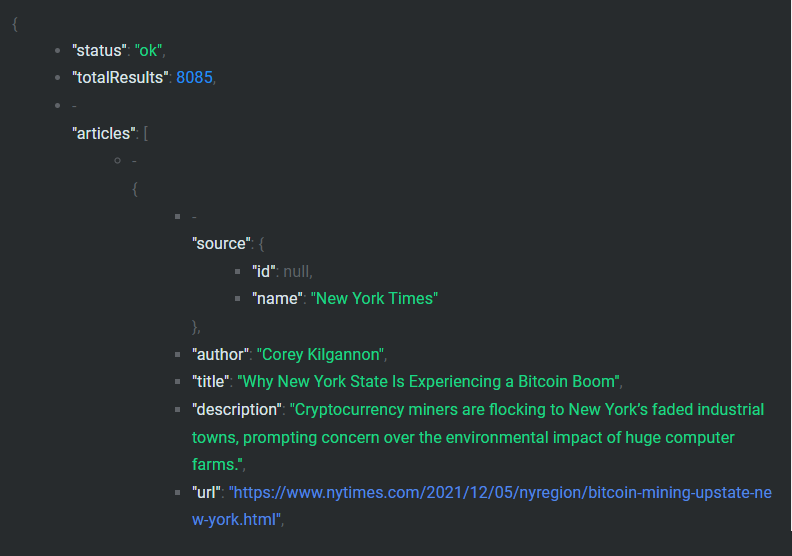


Fig 7: Example of News Data

**Sentiment Analysis**

It is also referred to as opinion mining, is an approach to natural language processing (NLP) that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service, or idea. In addition to identifying sentiment, opinion mining can extract the polarity (or the amount of positivity and negativity), subject and opinion holder within the text.

**Natural Language Processing**

Itrefers to the the branch of artificial intelligence or AI concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to ‘understand’ it’s full meaning, complete with the speaker or writer’s intent and sentiment.

CHAPTER 4  
 Experiment Analysis  
**4.1 System Configuration**

This project can run on commodity hardware. We ran entire projecton an Intel I5 processor with 8 GB Ram, 2 GB Nvidia Graphic Processor, It also has 2 cores which runs at 1.7 GHz, 2.1 GHz respectively. First part of the is training phase which takes 10-15 mins of time and the second part is testing part which only takes few seconds to make predictions and calculate accuracy.

**4.1.1 Hardware Requirements:**

• RAM: 4 GB

• Storage: 500 GB

• CPU: 2 GHz or faster

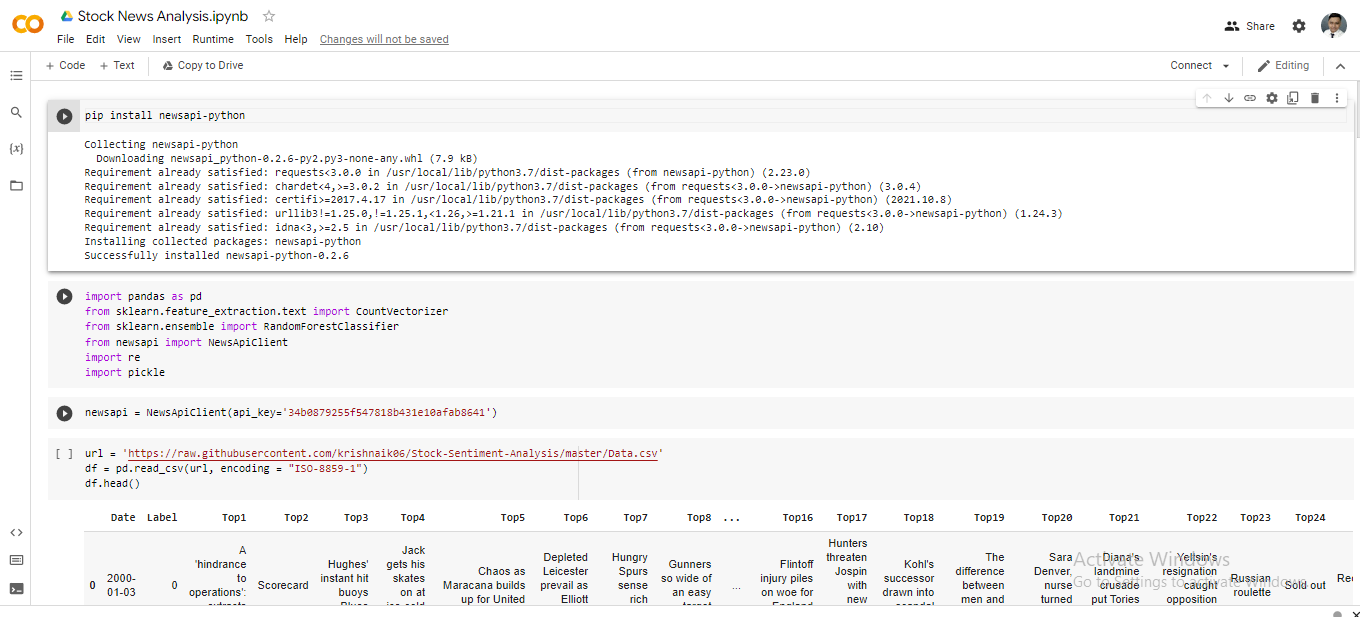
• Architecture: 32-bit or 64-bit

**4.1.2 Software requirements**

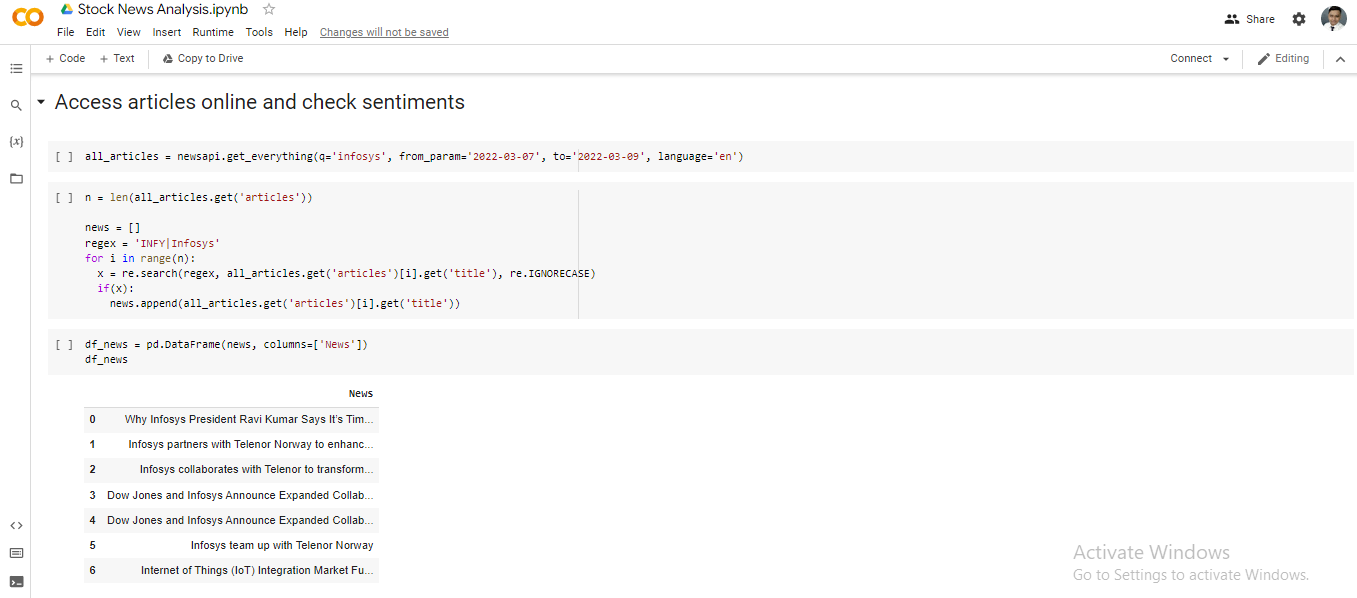
• Python 3.5 in Google Colab is used for data pre-processing, model training and

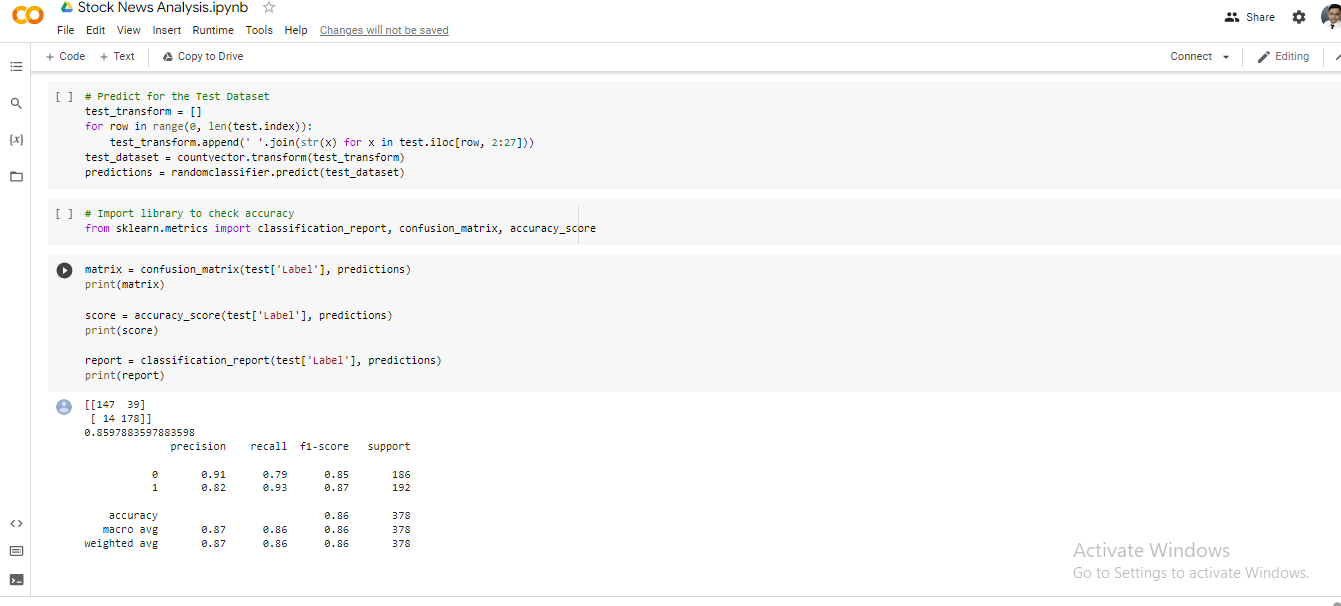
prediction.

• Operating System: windows 7 and above or Linux based OS or MAC OS.

**4.2 Sample Code :**





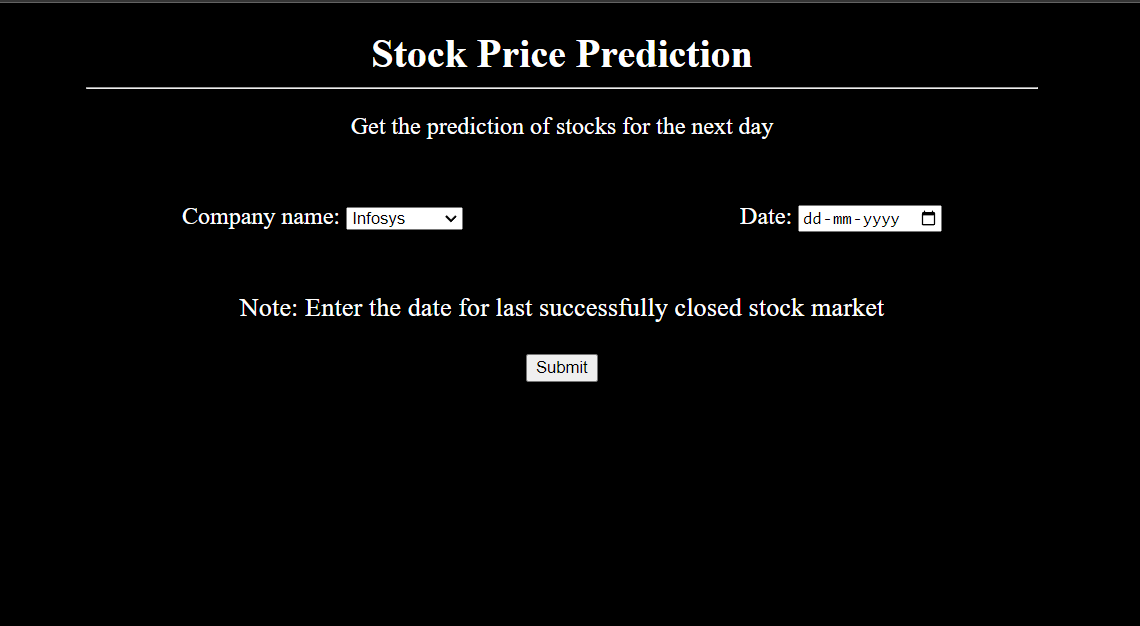




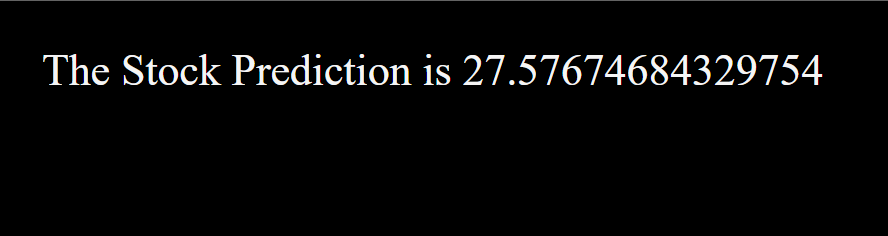


4.3 Sample Output:

Here we have to select the company name and date for last closed stock market .



Result of prediction on the given date.

**References**

[1] S. A. R. Nai-Fu Chen and Richard Roll, Economic Forces and the Stock Market, The Journal of Business, vol. 59, no. 3, pp. 383–403, (1986). [Online]. Available: <http://www.jstor.org/stable/2352710>.

[2] E. F. Fama, Random Walks in Stock Market Prices, Financial Analysts Journal, vol. 51, no. 1, pp. 75–80, (1995). [Online]. Available: <http://www.jstor.org/stable/4479810>.

[3] S. J. Grossman and R. J. Shiller, The Determinants of the Variability of Stock Market Prices, National Bureau of Economic Research, Working Paper 564, October (1980) [Online]. Available: <http://www.nber.org/papers/w0564>.

[4] A. W. Lo and A. C. MacKinlay, Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test, Review of Financial Studies, vol. 1, no. 1, pp. 41–66, (1988).

[5] R. Ahuja, H. Rastogi, A. Choudhuri and B. Garg, Stock Market Forecast Using Sentiment Analysis, In 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1008–1010, March (2015).

[6] P. A. G. Xue Zhang and Hauke Fuehres, Predicting Stock Market Indicators through Twitter I Hope it is not as Bad as I Fear, Procedia – Social and behavioral Sciences, vol. 26, pp. 55–62, (2011).

[7] J. Bollen, H. Mao and X. Zeng, Twitter Mood Predicts the Stock Market, Journal of Computational Science, vol. 2, no. 1, pp. 1–8, (2011). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S187775031100007X.