**Abstract**

**IACIT2022 -STOCK MARKET PREDICTION USING MACHINE LEARNING**

**Sai Bhavya Reddy.T1,Manju More2,Sanjana.S3,Shekamoori Pavithra4,Shwetha.V5**

**1,3,4,5-Student,Department of Computer Science and Engineering,**

**2-Assistant Professor,Department of Computer Science and Engineering,**

**REVA University.**

**1,2Student,School of Computer Science and Engineering,-Professor,School of Computer Science and Engineering,Sanjana.S3-Student,School of Computer Science and Engineering,Shekamoori Pavithra4-Student,School of Computer Science and Engineering,Shwetha.V5-Student,School of Computer Science and Engineering.**

Stock Market is where buying and selling of shares happens for publicly listed companies. The Stocks of the company which are also known as equities represent a fraction of ownership in companies. In the present finance world stock trading has very much importance. Buying and selling of shares happen through stock exchange. The oldest stock exchange is the Amsterdam Stock exchange, established in 1602. The Newyork Stock exchange, Chicago Board Operations exchange, American Stock exchange are the top exchanges in United States. we have the Bombay Stock Exchange(BSE) in India that was established in the year 1875 and it is Asia’s first stock exchange. We also have the National Stock exchange or NSE[1].

Stock Market Prediction is the process in which the future value of the financial stocks of a company are predicted. Now, Almost everywhere Machine Learning Techniques are used to predict the value of stocks. Machine learning techniques make predictions accurately taking values of current stock market indices on training the machine with old records. There are many methods in Machine learning which make predictions easier[2].

**Introduction**

In present times, it is difficult to analyze and predict the value of stock markets. dark pools, high-frequency traders, broker-dealers and the alternative trading systems always have exchanges among them. All of them interact with each other countless number of times in a fraction of second at every instant of a day[3].

It is also the result of Regulation National Market System partly which is also known as REG NMS.It is a set of rules passed to refine how all listed U.S. stocks are, by the securities and exchange commission(SEC) in 2005[12]. Reg NMS made in a way that no single body(may be an exchange), broker-dealer, or anyone else could have disproportionate control over the trading activity. The fragmentation that exists today is created by the decentralization of the market[8].

Tim Quast the founder and CEO of the Modern IR and Market Structure Edge has said that Regulation National Market System wanted businesses that were distinct ,independent to connect and share prices and customers, and to make that system work needs an immense amount of data technology and complexity and where all the complexity comes from and what has arisen out of that is what the businesses that comprise this ecosystem system have done to adapt to that[8].

Factors like Physical and Psychological factors, rational and irrational behavior etc affect values of stock market. Also there are factors like market volatility, dependent, independent to decide value of a stock in the market. It very difficult to predict the value of stocks based on these factors for any stock market analyst with high accuracy. But, Machine Learning helps in predicting the stock market value by understanding the stock market data through its techniques[9].

With an idea of stock and the stock marketing, let’s understand why people believe that by a Machine Learning model they can predict the stock price. The predictions of values of an observation by a machine learning model are dependent on several inputs that are predictors. the value of an observation based on several inputs that are predictors. The stock market is working similarly, that is, based on several inputs, the stock price fluctuates with these factors. So machine learning has to keep all these factors included and predict the shares of the stock market.

Machine Learning at its fundamental level requires the user to provide the previously existing data using this previous data, the machine learning techniques get trained and will be able to analyze and predict the future values of new data. In this paper, we will use machine learning techniques like linear regression and long short term memory network to predict the future values of stock markets[4].

Linear Regression is a supervised learning algorithm to predict the outcome of a continuous variable. It is a very popular statistical technique to solve machine learning problems.it can be used to predict total revenue of the company, weather prediction, Stock price prediction etc.[10]

Long Short term memory(LSTM) is a type of recurrent neural network for learning long-term dependencies. It is commonly used for processing and predicting on the basis of time series data. LSTMs have a chain like structure[5].

**METHODOLOGY**

It is very difficult and complex to predict stock market as there are so many factors that it depends on. In this paper, we use machine learning techniques called linear regression algorithm and long short term memory network as discussed above by relating the previous data to the current data. we train the machine with a set of data so that it will be able to analyze and predict any given test data[2].

**1.Regression Based Model**

It is used to predict the outcome of a continuous variable. It is based upon the use of regression algorithm for predicting correct values. The factors that are considered here for regression are date, open, high, low, close, adjusted closing price. Here, date is the price at which the stock started trading when the market opened on a particular date close is the price of an individual stock when the stock exchange closed the market for the day. It represents the last buy-sell order executed between two traders. High is the highest price at which a stock traded during a period. Low is the lowest price of the period. Volume here is the total amount of trading activity that took place during certain period of time. Adjusted closing price is calculation adjustment made to the stocks closing price, more complex and accurate than the closing price. The adjustments made to the closing price depicts the true price of the stock because the outside factors would have altered the true price[2]

The work is carried out on csv format of data through libraries like pandas, numpy and matplotlib etc in jupyter notebook. Linear regression is made on the data and predictions are made. Regression uses a linear function previously given for predicting continuous values[2]

Pre classified output

shares

Test Data

Training data

Algorithm in machine learning

Regression based model

Predicted Output

**Figure1:Regression Based Model**

Regression based model is generally used to predict the continuous values using some given independent values. Figure-1 minimizes the error function and predicts the given values. Regression generally uses the given linear function to predict the continuous values.

V = a + bK + error

In the above function, V represents a continuous value; K represents known independent values and a,b are the coefficients[2].

**2.Long Short Term Memory Model:**

Long Short Term Memory Modelis a type of recurrent neural network to learn long-term dependencies. Long Short Term Memory Model is commonly used to process and predict on the basis of time series data. LSTMs have a chain like structure.[11]

Generally, in this method first step is used to decide on the information that is to be excluded from the cell in that particular time. Sigmoid function helps in deciding it. It takes the previous state and current input state into consideration and computes the function. There are two parts in the second step, Sigmoid function and the tanh function. The sigmoid function decides on which values to let through. In the third step, the final output is decided. A sigmoid layer is executed, which decides the parts of the cell state those make it to the output and then we put the cell state to the tangent function and multiply it by sigmoid.

Long Short Term Memory(LSTM) is a well versed version of Recurrent neutral networks(RNN). Long Short Term Memory and Recurrent neutral networks are not same as LSTM is the advanced version or different version of RNN. Long Short Term Memory(LSTM) involve long term dependencies. Recurrent neural networks(RNN) work to find the relationship between the current and recent information.

The problems of conventional recurrent neural network (RNN) which can be the vanishing and exploding gradients can be provided with a solution to these problems in the form of Long Short Term Memory (LSTM) Long Short-Term Memory is the well versed version of (RNN) architecture. It is used designed to model chronological sequences and their long-range dependencies more precisely than conventional RNNs. In many situations a reference to certain information stored quite a long time ago is required to predict the current output. But “long-term dependencies” cannot be handled by RNNs. Here there is no finer control over which part of the context needs to be carried forward and how much of the past needs to be ‘forgotten’. Thus here the LSTM is used where the vanishing gradient problem is almost completely removed, while the training model is left unaltered. Long time lags in certain problems are solved using LSTMs where they also handle noise, distributed representations, and continuous values. The basic difference between the RNNs and LSTMs is in the architecture where the hidden layer of LSTM is a gated cell. LSTMs consists of 3 logistic sigmoid gates and one tanh layer. These gates have been introduced to limit the information that is passed through the cell.

LSTMs also make use of input and forget gate instead of two separate gates that helped in making both the decisions simultaneously. Deep LSTM will involve a number of LSTM layers in between the input and output.  Deep LSTM with a Recurrent Projection Layer will have numerous LSTM layers where each layer has its own projection layer. In the case where the memory size is too large increased depth is very usefull. It has certain layers like –

layer=lstmLayer(numHiddenUnits) creates an LSTM layer and Num hidden  property.

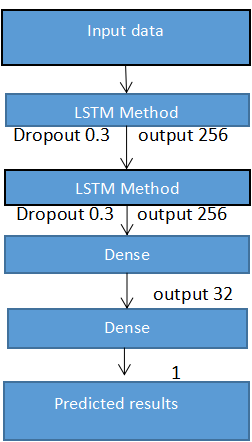
layer=lstmLayer(numHiddenUnits, Name, Value)sets additional Output activities, parameters and initialization, name properties using one or more name-value pair arguments. You can specify multiple name-value pair arguments, by enclosing each property name in quotes.

Here at first, the output of an LSTM at a particular point of time is dependent on three things –

1.The forget gate is the first step in the process. Here we will decide which bits of the cell state (long term memory of the network) .

2.The new memory network and the input gate is involved in this step. The goal of this step is to determine what new information should be added to the networks long-term memory .

3.One of the person might think that we can just output the updated cell state; but however, this would be comparable to someone else who is unloading everything they learned about the stock market when only asked if they think it will go up or down tomorrow. To prevent this from happening we create a filter, the output gate, exactly as we did in the forget gate network[13].

****

**Figure 2- LSTM Layers**

Figure2 indicatesthat the interval of information is relatively smaller compared to LSTM. The main reason to use this model in stock market prediction is due to that the predictions depend on large amounts of data and are generally dependent on the history of the stock market.[2]

**Result:**

The above systems are trained and tested over the dataset taken from Tesla for first method and google for second method.It is split into to training set and testing set respectively. The following results are yielded upon passing through the different models.

1. **Regression Based Model Results**

The following graph is obtained on applying the linear regression algorithm on the dataset of tesla from 29-06-2010 and 15-03-2019 to predict varying prices.

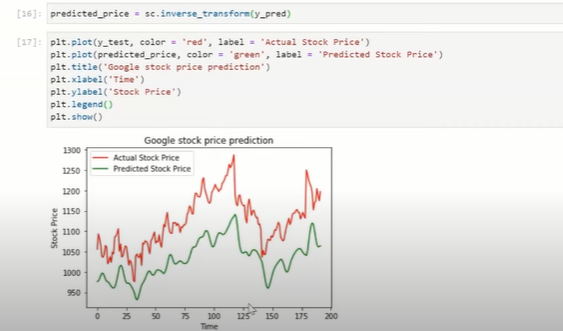


**Figure-3:Stock Prices of Tesla**

The actual stock prices are shown in blue and the predicted ones are shown by red line

1. **LSTM Based Model Results**

The following graph is obtained on applying the linear regression algorithm on the dataset of Google from 13-08-2018 and 13-08-2019 to predict varying prices.



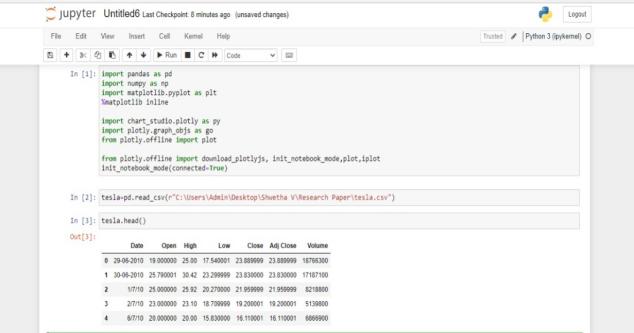
**Figure-4:Google Stock Price Prediction**

Red indicates actual stock price and green indicates predicted stock price.

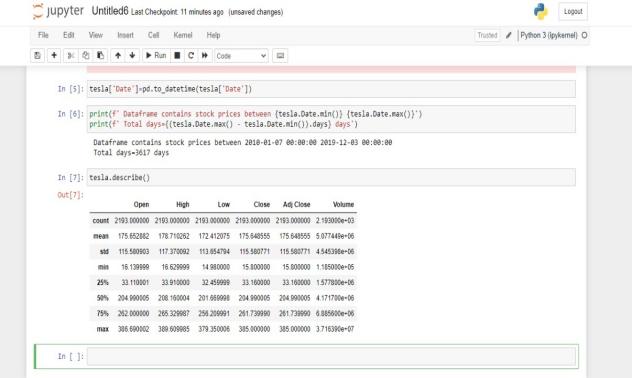
**Implementation:**

Following are some pictures of implementation with inputs of Tesla and Google companies and respective outputs

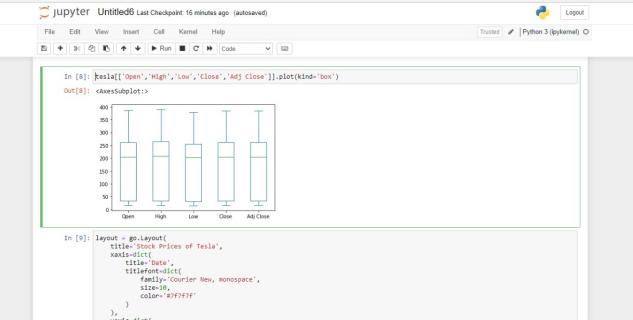
1. **Regression Based Model:**



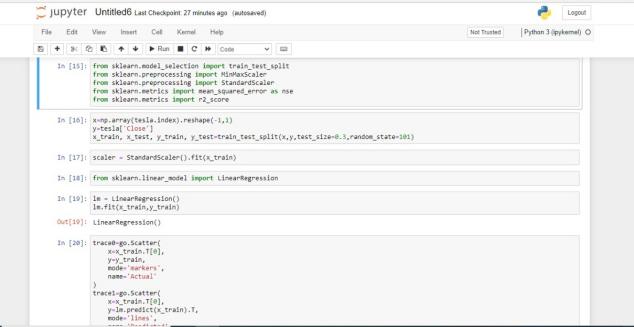
**Figure-5: Top 5 rows from the dataset.**

****

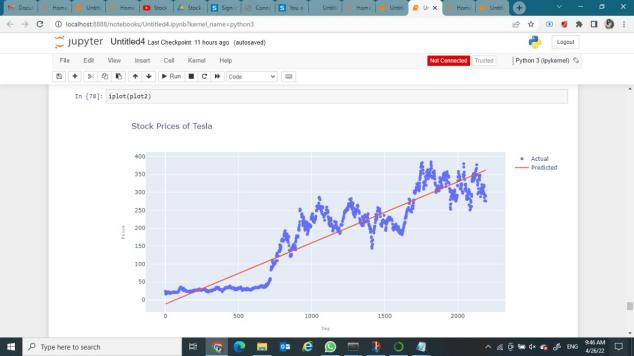
**Figure6:some summary statistics for the data**

****

**Figure7: Boxplot**

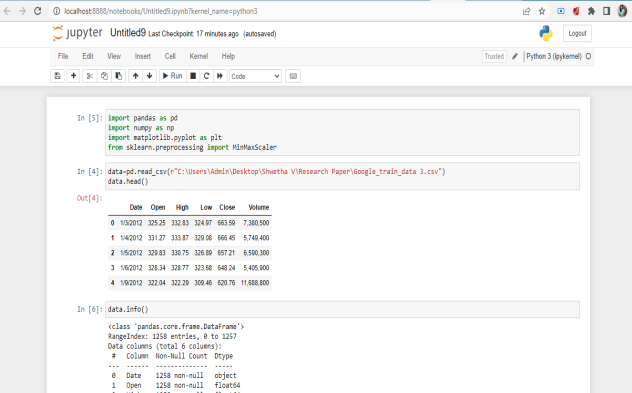


**Figure-8:This part gives the graph which contains actual and predicted stock values using iplot function.**

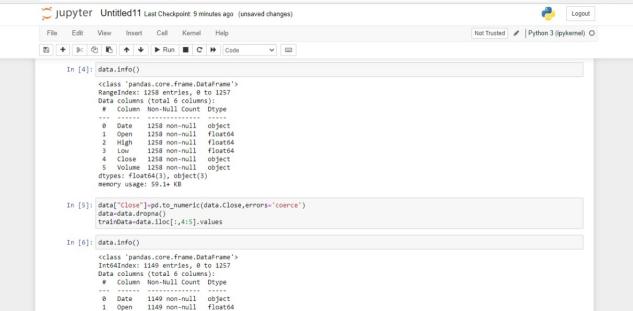


**Figure-9:Actual and Predicted Stock Prices of Tesla**

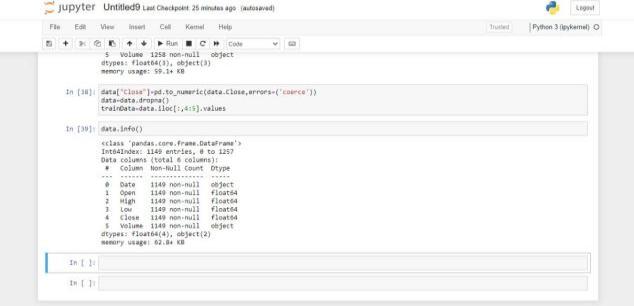
1. **LSTM Based Model:**



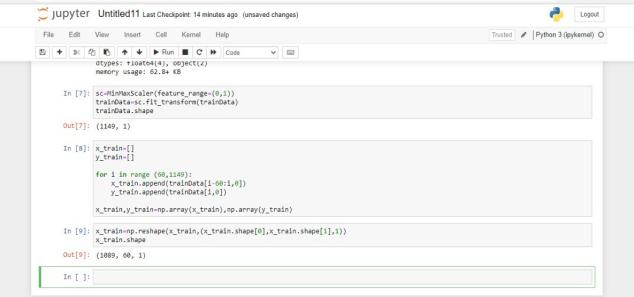
**Figure-10: Top 5 rows from the dataset.**



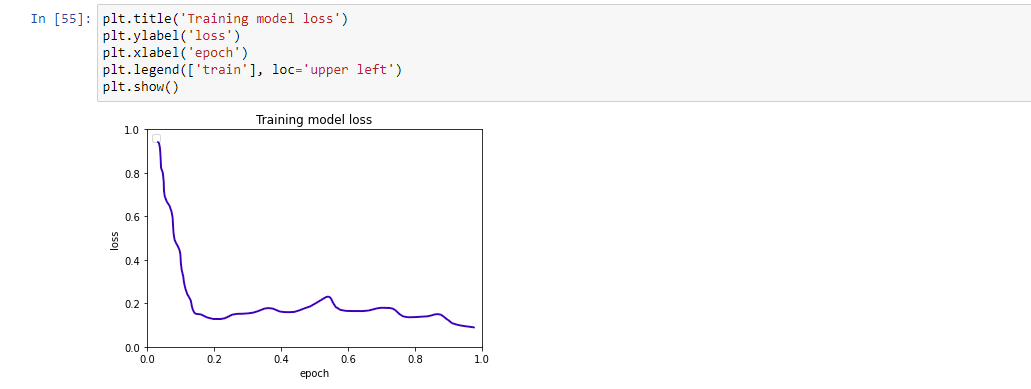
**Figure-11: total no of rows and columns in the dataset and the data types of each variable.**



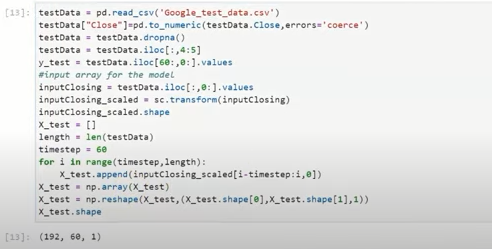
**Figure-12:change of datatype of closed column from object to float**



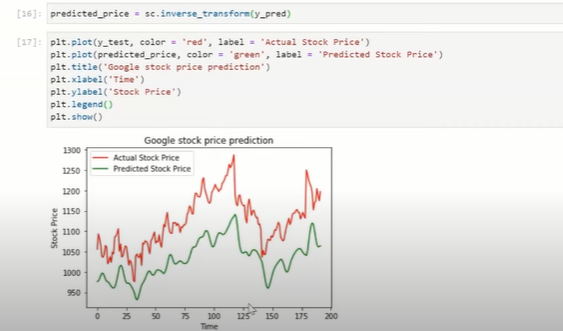
**Figure-13: Reskilling the data between 0 and 1 for better performances.**



**Figure-14:visualization of loss that occured during training process for each epoch**



**Figure-15:Testing of model with a new dataset**



**Figure-16:Actual and Predicted Stock values of Google.**

**Conclusion**

Investors commonly predict the stock prices to know the amount they get back. Generally,Stock prices were predicted in traditional way by Brokers and Technical Analysts based on previous prices,volumes,price patterns and the basic trends. But, present it has become very difficult and complex task to predict stock markets because now stock markets are dependent on many other factors like social and economical conditions of the country ,natural disasters and political atmosphere etc.Sometimes,The returns that investors get back is very uncertain and disastrous and is difficult to predict using traditional methods.So, a lot of research is made on finding the methods that can predict the accurate values of stocks .Some of them are machine learning techniques like LSTM AND Regression Based Model.These kind of methods help Stock Brokers and Finance Institutions in getting good return [6].

**Future Work**

In the future,a much bigger dataset can be used than the one being used currently thus increasing the stock market prediction system. This leads to increase in the accuracy of our prediction models.Further,to predict the values of stock prices other models of Machine Learning can also be studied[7].

**References**

1. “Stock Market Prediction using Machine Learning” V Kranthi Sai Reddy-Student,ECM,Sreenidhi Institute of Science and Technology,Hyderabad,India.
2. “Stock Market Prediction Using Machine Learning” Ishitha Parmar , Navanshu Agarwal , Sheirsh Saxena , Ridam Arora, Shikhin Gupta , Himanshu Dhiman , Lokesh Chouhan- Department of CSE -NIT Hamirpur,India.
3. M.Usmani, S.H. Adil, K.Raza and S.S.A. Ali,”Stock market prediction using machine learning techniques”,2016 3rd International Conference on Computer and Information Sciences( ICCOINS),KualaLumpur,2016,pp.322-327.
4. “LSTM and Regression Methods for Stock Market Prediction”-Nishanth Vaidya,Nikhil Bharadwaj-Students,Department of CSE-Sambhram Institute of Technology,Bangalore,India.
5. “Stock Closing Price Prediction using Machine Learning Techniques”-Mehar Vijh,Deeksha Chandola,Vinay Anand Tikkiwal,Arun Kumar-Jaypee Institute Of Information Technology,India.
6. H.L. siew and M.J. Nordin,”Regressioon techniques for the prediction of stock price trend “,2012 international conference on statistics in science,Business and Engineering(ICSSBE),Langkawi,2012,pp. 1-5.
7. “Stock Market Prediction Using Machine Learning Algorithm” Sayyed Aman,Sayyed Gulfam,Shaikh Naba-UG Students-Mumbai University-Mumbai,India.
8. [**https://www.analytixlabs.co.in/blog/stock-market-prediction-using-machine-**](https://www.analytixlabs.co.in/blog/stock-market-prediction-using-machine-learning/)**learning/** [9][**https://www.adityabirlacapital.com/abc-of-money/factors-**](https://www.adityabirlacapital.com/abc-of-money/factors-affecting-stock-market)**affecting-stock-market**

**[10]**[**https://machinelearningmastery.com/linear-regression-for-machine-**](https://machinelearningmastery.com/linear-regression-for-machine-learning/)**learning/**

**[11]**[**https://en.m.wikepedia.org/wiki/Long\_short-term**](https://en.m.wikipedia.org/wiki/Long_short-term_memory)**\_memory**

**[12]** [**https://www.investopedia.com**](https://www.investopedia.com)

**[13]**<https://www.geeksforgeeks.org/understanding-of-lstm-networks/>