ANALYSIS OF STOCK PRICE PREDICTION USING MACHINE LEARNING APPROACH

[1]Nikhil Khandare, [2]Harshali Patil
[1]Student, [2]Associate Professor
[1]MET Institute of Computer Science, Bandra(W), Mumbai, India

Abstract: Nowadays the most difficult task regarding stocks is to predict them. There are various factors used to predict stock prices. Each of these perspectives integrates to make share costs unpredictable and hard to anticipate with a serious level of exactness. Great and successful forecast models for stock prediction help dealers, financial backers, and analysts by giving steady data like the future value of the Stock Market. This paper incorporates Predicting stock costs using LSTM (Long Short-Term memory), Linear Regression and Moving Average techniques. Accuracy of all three algorithms is compared using RMSE values.

Keywords **- stock prediction, moving average, linear regression, LSTM.**

I. INTRODUCTION:

Each business seeks to grow. But only a handful of organizations effectively realize the vision to do so through data-based decision making. organizations have been utilizing machine learning-based predictive analytics for the same. Predictive analytics is predicting future results dependent on current and historical data. It uses different statistical and data modeling strategies to examine the past data, distinguish patterns, and helps to make business decisions.
The financial market is where the stocks are moved, exchanged, and circulated. The stock market prediction is the demonstration of trying to decide the approximate prediction of an organization's stock or other budgetary related instrument traded on a financial exchange. The fruitful expectation of a stock's future cost will amplify the financial backer's benefits.
This paper proposes a technique of predicting stock price using the LSTM, Linear Regression and Moving average algorithms of Machine Learning. It is trained using historical data and real-time data to lessen the risk of the investor and maximize the profit of the company. The linear regression method is used to model the relationship between a scalar dependent variable y and one or more explanatory (or independent variables) denoted X [1]. LSTM is extremely incredible in sequence prediction problems because they are able to store preceding data. This is significant for our situation in taking note of the fact that the past cost of a stock price is vital in forecasting its future cost [2]. One of many time series analysis techniques is the simple moving average. Time series analysis is a method of processing structured data in real time in order to find statistics or important characteristics for a variety of reasons [3].

II. LITERATURE REVIEW

In this research paper [1], artificial neural networks are utilized as a model which predicts stock trends using Simple moving average and financial news sentiment. The stocks are affected by many factors, one of which is that daily news, especially financial news, plays a major role in predicting stock price. Every paper has different sentiment values like positive, negative and neural sentiment that directly affects the stock trend whether it will go up or down. The technique utilized is restricted to make predictions. It would be more intriguing to utilize prediction techniques like Regression and LSTM.
This paper [2] implements well-known efficient regression approaches to predict stock price. Polynomial regression, Rbf regression, sigmoid regression and linear regression are implemented to predict stock price. The results of this model could be better using more variables.

The following paper [3] uses neural network techniques to predict stock price. Recurrent neural network (RNN) with memory capability is implemented. Long Term Short Memory (LSTM) architecture is used in this paper to improve its performance. More historical data can improve the model accuracy

III.PREDICTION METHOD:

3.1 LINEAR REGRESSION:

Linear regression is the most basic machine learning technique to predict the stock price or trend. Linear regression provides an equation where we have independent variables which are our features and dependent variable which are the value we are predicting [Stock price in our case].
Linear regression equation:

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots \theta_n X_n$$

Here, X's are the independent variables, Y is the dependent variable (Stock Prices) and thetas are the coefficients. Coefficients are the weights assigned to the independent factors, based on their significance.

3.2 MOVING AVERAGE:

'Average' is very commonly used in day-to-day lives. For example, to calculate today's temperature we can average the temperature of the past few days or to calculate the overall performance we can calculate the average marks. The predicted closing price of the stock for each day will be the average of previously observed closing prices. Rather than using a simple average, we will utilize the moving average technique which uses the most recent set of values for each prediction. All in all, for every following step, the predicted values are considered and oldest observed values are removed.

Following diagram is the explanation of the Moving average method:

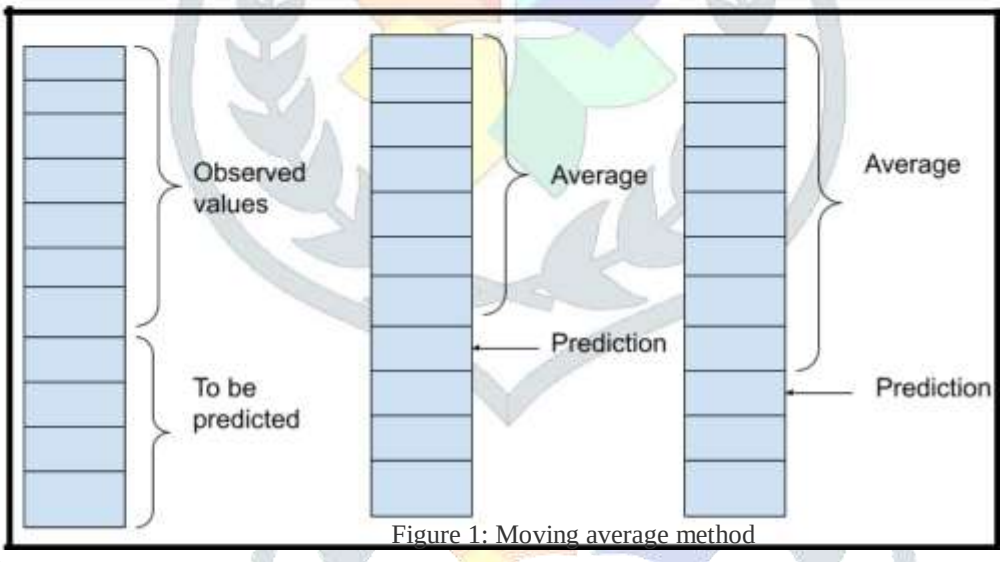

Figure 1: Moving average method

3.3 LSTM:

LSTM are generally utilized for sequence prediction problems and have demonstrated to be amazingly effective. The basis they work efficiently is because LSTM can store the past data that is significant and ignores the data that is not.
The following are the three gates of LSTM:
The input gate: Responsibility of the input gate is adding data to the cell state.
The forget gate: Responsibility of the forget gate is to eliminate the information that is not required by the model anymore.
The output gate: Responsibility of the output gate is to choose the data to be shown as output.

The LSTM is a form of RNN's. RNN's measures input in a sequential way, where the data from the past input is involved to predict the output of the current step. This permits the neural network to convey data throughout various time steps as opposed to keeping each one of the information sources free of one another. RNN's can't remember the data from much earlier whereas LSTMs can hold onto data for the long-term.

IV.IMPLEMENTATION:

Initially, we used linear Regression Algorithm for predictive analysis and afterwards different machine learning algorithms are compared on the basis of RMSE (Root Mean Square Error).

4.1 Predictive Analysis

In this research paper, we have collected the stock data of apple company from Kaggle. Predictive analysis is the effort to predict what may happen. To do this, we need to dig into data and pull out the portions that could assist in solving that question. After this, the gathered data was manipulated and undesirable data was eliminated in excel and it was saved in the form of a CSV (Comma Separated Value) document because the model is constructed in a Jupyter notebook (Python). Later the dataset was imported into the jupyter notebook using pandas which is a python library which is a high-performance data analysis tool. The dataset is shown in the figure:

	Date	Open	High	Low	Close	Adj Close	Volume
0	2015-05-18	18.000000	18.000000	17.030001	18.000000	12.911796	2197500
1	2015-05-19	17.920000	19.150000	17.920000	18.799999	13.485645	1978700
2	2015-05-20	18.700001	18.840000	18.360001	18.610001	13.349364	717600
3	2015-05-21	18.600000	18.660000	18.020000	18.250000	13.091125	793000
4	2015-05-22	18.250000	18.400000	18.049999	18.209999	13.062433	536200
5	2015-05-26	18.250000	18.450001	18.110001	18.240000	13.083947	681400
6	2015-05-27	18.200001	18.700001	18.000000	18.650000	13.451800	652700
7	2015-05-28	18.430000	18.700001	18.320000	18.520000	13.358037	949800
8	2015-05-29	18.330000	18.500000	18.110001	18.450001	13.307544	565500
9	2015-06-01	18.330000	18.500000	18.250000	18.480000	13.329185	1014500

Figure 2: Dataset in jupyter notebook

Dataset contains following fields:
Date: The trading day's date
Open: On Date, the first trade price was set.
High: On Date, the highest price at which the stock is traded.
Low: On Date, the lowest price at which the stock is traded.
Close: Date's most recent trade price
Adj Close: This is the closing price after all dividends have been distributed.
Volume: The total number of shares traded on the given date.

The Date field was converted to integer data format using the datetime library for linear regression. Afterward, a train-test split was utilized to part the dataset into training data and testing data in the proportion of 80:20 separately. The training data accommodate known output and it is used further to train the model. Test data is used to test the prediction of the model. Finally, machine learning models were executed to predict the stock values.

4.2 Comparison:

In this research paper, distinctive Machine Learning Algorithms are contrasted with examination of which one gives the best outcomes and these calculations are looked at based on RMSE (Root Mean Square Error) Value.

V. RESULT ANALYSIS:

Various Machine learning algorithms are compared**,** and the results are shown in the table and graphical format.

Table 1: Comparison of the experimental results of the three algorithms.

SR.NO.	ALGORITHMS	RMSE VALUE
1.	LR	1.2
2.	MA	7.6
3.	LSTM	1.3

Where,
LR is a Linear Regression.
MA is a moving-average.
LSTM is a Long short-term memory**.**

As the table and graph shows that Linear Regression has the most reduced RMSE demonstrates the best model when contrasted with others for the same data.

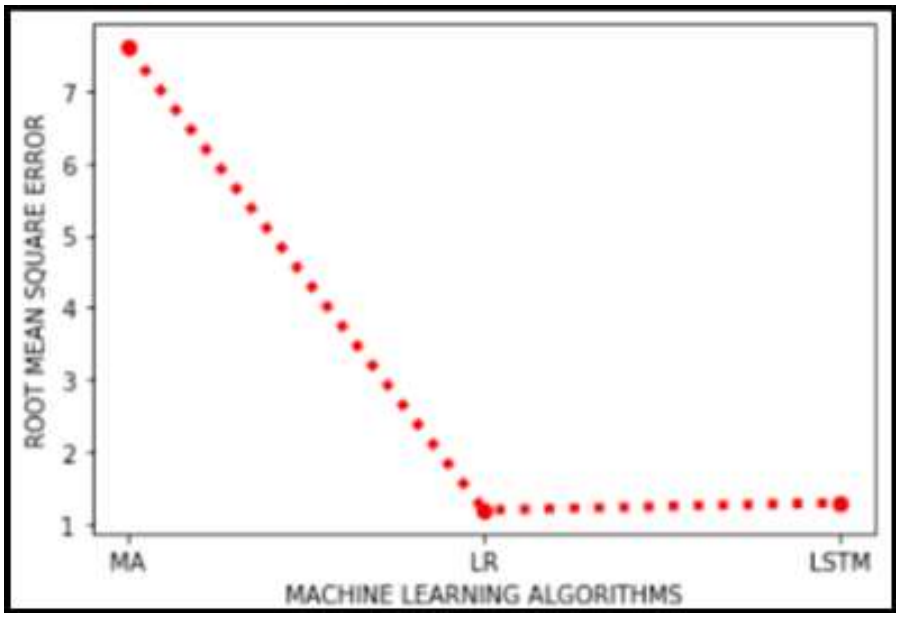

Figure 3: Machine Learning model accuracy comparison

As the RMSE value of the linear regression model is better compared to other two algorithms we have calculated the r2 score of the regression model:

```
In [136]: from sklearn.metrics import r2_score
          score = r2_score(Y_test, pred, sample_weight=None)
          print("Score:",score)

          Score: 0.4919640395590291
```

Figure 4: Score of Linear regression model

Heatmap of the dataset:

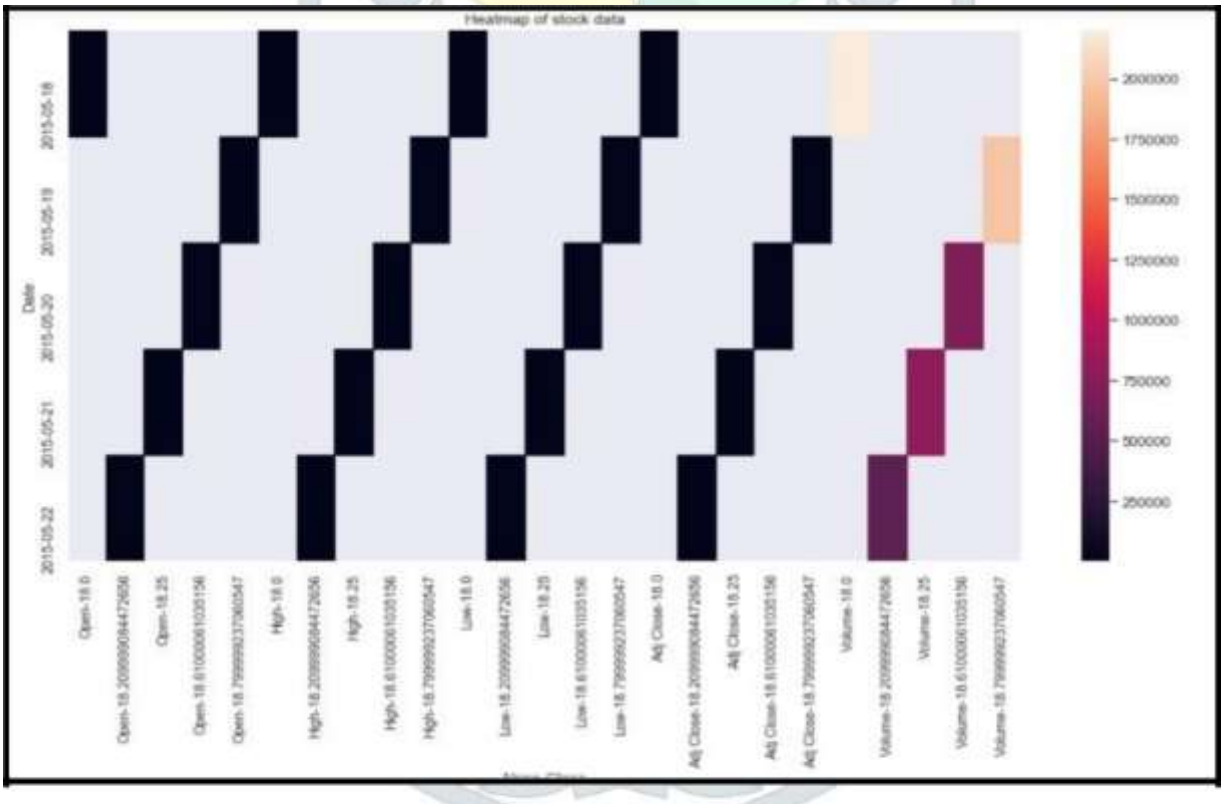

Figure 5: Heatmap

VI. CONCLUSION:

Therefore, in this paper the main goal of predictive analytics is to predict stock price during execution. We also compared the RMSE (Root Mean Square Error) values of various algorithms to verify the accuracy of the model and concluded that the Linear Regression has the lowest RMSE value.

REFERENCES

[1] ASHISH SHARMA, DINESH BHURIYA, UPENDRA SINGH, "Survey of Stock Market Prediction Using Machine Learning Approach", Govt. Women's Polytechnic, Indore, 2017

[2] Bo-Sheng Lin, Wei-Tao Chu, Chuin-Mu Wang, "Application of Stock Analysis Using Deep Learning",Department of Computer Science and Information Engineering National Chin-Yi University of Technology Taichung, Taiwan (R.O.C.), 2018.

[3] Stefan Lauren, Dra. Harlili S., M.Sc., "Stock Trend Prediction Using Simple Moving Average Supported by News Classification", Informatics Engineering Bandung Institute of Technology Bandung, Indonesia, 2014.