Grenze International Journal of Engineering and Technology, Jan Issue

Stock Market Prediction Analysis using Machine Learning

Rachit Kothari[1], Khushank Mehta[2], Sarvesh Naik[3], Parth Powale[4] and Mamta Tikaria[5]
[1-4]Shah and Anchor Kutchhi Engineering College, Student of EXTC Department, Mumbai, India
Email: { rachit.kothari, khushank.mehta, sarvesh.naik, parth.powale}@sakec.ac.in
[5]Shah and Anchor Kutchhi Engineering College, Asst. prof. EXTC Department, Mumbai, India
Email: mamta.tikaria@sakec.ac.in

***Abstract*—In Stock Market Prediction, the aim is to predict the future value of the financial stocks of a company. Machine learning based predictions are very popular now these days which makes predictions based on the values of current stock market indices by training on their previous values. Machine learning employs different models to make prediction easier and authentic. This paper focuses on the use of Long Short-Term Memory (LSTM) based Machine learning to predict stock values. Factors considered for accurate prediction are open, close, low, high and volume.**

***Index Terms*— Stock Market, financial stocks, LSTM, Machine Learning, Prediction, training, models.**

I. INTRODUCTION

The stock market appears in the news every day. One can hear about it every time it reaches a new high or a new low. The rate of investment and business opportunities in the Stock market can increase if an efficient algorithm could be devised to predict the short-term price of an individual stock [1]. Investors are familiar with the saying, "buy low, sell high" but this does not provide enough context to make proper investment decisions. Before an investor invests in any stock, he needs to be aware how the stock market behaves. Financial investors of today are facing this problem of trading as they do not properly understand as to which stocks to buy and sell for optimum profits. An accurate prediction of future prices may lead to a higher yield of profit for investors through stock investments. As per the predictions, investors will be able to pick the stocks that may give a higher return.

II. LSTM MODEL

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. LSTM's have an edge over conventional feed-forward neural networks and RNN in many ways. This is because of their property of selectively remembering patterns for long duration of time. LSTM networks are well-suited to classifying, processing and making predictions based on time series data.

LSTM has a chain structure that contains four neural networks and different block called cells.

1. Forget Gate: The information that is no longer useful in the cell state is removed with the forget gate. Two inputs x_t(Input at particular time) and h_t-1(previous cell output) are fed to the gate and

Grenze ID: 01.GIJET.9.1.103

multiplied with the weight matrices followed by addition of bias. The resultant is a particular cell state the output is 0, the piece of information is forgotten and for output 1, the information is retained for future use.

2. Input gate: The addition of useful information to the cell state is done by the input gate. First, the information is regulated using the sigmoid function and filter the values to be remembered similar to the forget gate using inputs h_t-1 and x_t. Then, a vector is created using tanh function that gives an output from -1 to +1, which contains all the possible values from h_t-1 and x_t. At last, the values of the vector and the regulated values are multiplied to obtain the useful information.
3. Output gate: The task of extracting useful information from the current cell state to be presented as output is done by the output gate. First, a vector is generated by applying tanh function on the cell. Then, the information is regulated using the sigmoid function and filter by the values to be remembered using inputs h_t-1 and x_t. At last, the values of the vector and the regulated values are multiplied to be sent as an output and input to the next cell.[2]

III. Proposed Work

This paper used historical data about the stock prices of a publicly listed company and implement the machine learning algorithm to predict the future stock price of any company, using the LSTM algorithm and keras model using python.

Figure 1. Block diagram of proposed work

IV. Stock Market Prediction Using LSTM

Long-Short-Term Memory (LSTM) Recurrent Neural Network is one of the popular deep learning models, used in stock market prediction. In this task, one can fetch the historical data of stock automatically using python libraries and fit the LSTM model on this data to predict the future prices of the stock. In this task, the future stock prices of State Bank of India (SBIN), India Tobacco Company Limited (ITC), ZEE Entertainment, Hindustan Uniliver (HUL) and Apple are predicted using the LSTM Recurrent Neural Network. Here task is to predict stock prices for a few days, which is a time series problem. The LSTM model is very popular in time-series forecasting, and this is the reason why this model is chosen in this task. The historical prices of SBIN, ITC, ZEE, HUL and APPLE are collected automatically using the pandas library of python by using the pandas data reader. This paper used 8 years of historical price data, from 01.01.2012 to 04.01.2021.

V. Working

This program uses the Artificial Neural Network called Long Short-Term Memory (LSTM) to predict the future stock price of the company. LSTM is used widely for sequence prediction problems. The steps involved are-
a) Importing Libraries: We import various libraries such as math library, pandas_datareader as web, numpy, pandas, Min Max Scaler from sklearn. Preprocessing, Sequential from keras. model, Dense, LSTM from keras. layers and matplotlib library.

b) Get the Stock Quote: Use the web. Datareader to get the data of the specific stock from a particular start date till a particular end date. In the data we see that there are some missing values, these missing values are the one on which the market was close due to weekends or public holidays.
c) Get the Number of rows and columns in Data Set: This paper used the df.shape in order to view the rows and columns in the dataset.
d) Visualize the Closing Price of Stock: This paper used the matplotlib library in order to visualize the closing price of the data by giving the name of the stock we want and providing the unit for the currency such as rupees dollar etc.
e) Create a new Dataframe with only the Close column: Create a new dataframe using the pandas library and then convert it to numpy array for the data processing.
f) Scaling the Data: Before feed the data to any neural network first scale the data in order to gain good results. This paper used the MinMaxScaler in order to scale our data by providing the feature range of (0,1). The scaler.fit_transform computes the minnimum and maximum values to be used for scaling and transforms the data into these values. We have provided the value (0,1) so we will get any value between (0,1) both inclusive.
g) Create the training data set: Before creating the training data first scale the data then split the data into x_train and y_train. X_train is independent training variable and y_train is the dependent training variable.
h) Appending the data: Append the last 60 days data to the x_train data set. The y_train data will contain the 61st data value.
i) Convert the x_train and y_train to numpy arrays
j) Reshape the data: LSTM accepts the input to be in 3D therefore we reshape our 2D data into 3D data.
k) Build the LSTM Model: Create the LSTM layers using the kereas sequential model. This paper provides 50 neurons to first LSTM Layers and the return_sequence is kept True ass we are going to provide more number of layers to it. In the second LSTM Layer, again provide 50 neurons and the return sequence is kept False as not going to create more layers after it. Then add a Dense layer with 25 neurons which is just a basic densely connected layer. At last create a second dense layer with just 1 neuron.
l) Compiling the model: Compile the model using the adam optimizer to improve upon the loss function and the loss function is used to measure how well the model did on training.
m) Train the model: Fit the training model into the LSTM layers and provide a batch size and epoch.
n) Testing the Model: Create 2 Data Sets, x_test and y_test. The y_test is the values model want to predict. And the x_test data set is the past 60 days values which are using for testing purpose.
o) Get the model predicted price value: Once model is trained and give the testing data to it then the predicted value given by testing data which is the y_test. After that give the RMSE (Root Mean Square Error) value to evaluate the model.
p) Plotting the data: With the help of matplotlib library, plot the predicted and actual data in order to show the original and predicted value.

VI. RESULTS

The LSTM algorithm was used to predict the closing price of the Stocks. The RMSE (Root Mean Square Error) was found to be the least in all the three stocks. The actual value of the stock and the one we predicted were almost equal to each other. This shows that the LSTM gives better accurate performance. So, this technique will help the traders to minimize the loss and attain maximum profit. For prediction of closing price of all stocks given below, algorithm uses 10 years data which was collected from Yahoo Finance. Figure 2, figure 3, figure 4, figure 5 and figure 6 shows the precited stock values of SBI, APPLE, ITC, ZEE Ltd and Hindustan Unilever company respectively.
Table 1 provides the stock prediction of 50 companies across various fields and their predicted closing price compared with actual closing price dated 27th April 2022.From the research conducted on fifty stocks given below, the specified model has achieved total accuracy of 95% and the average difference is 5%.

VII. CONCLUSION

With the introduction of Machine Learning and its strong LSTM algorithm, the most recent market research and Stock Market Prediction advancements have begun to include such approaches in analyzing stock market data. The Opening Value of the stock, the Highest and Lowest values of that stock on the same days, as well as the Closing Value at the end of the day, are all indicated for each date. Furthermore, the total volume of the stocks in the market is provided, with this information, it is up to the job of a Machine Learning Data Scientist to look at

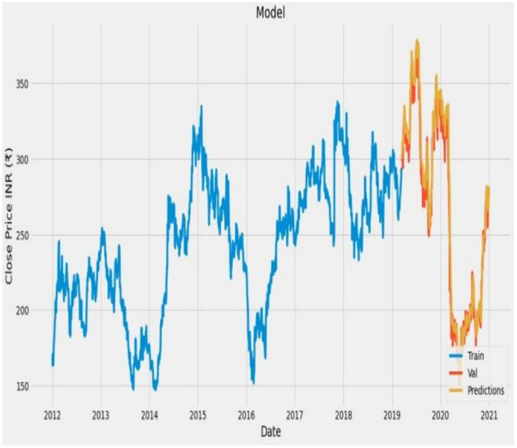

Figure 2. SBI Stock Prediction

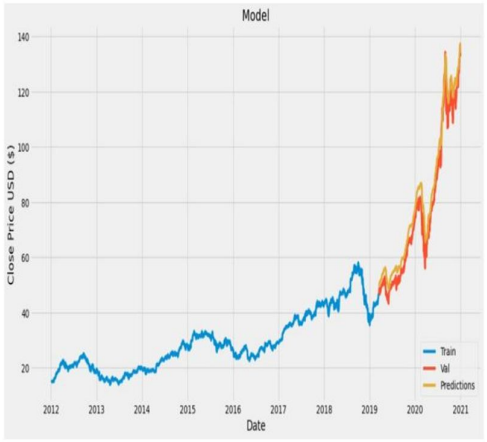

Figure 3. APPLE Stock Prediction

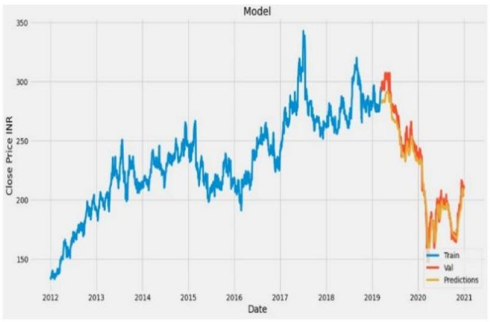

Figure 4. ITC Stock Prediction

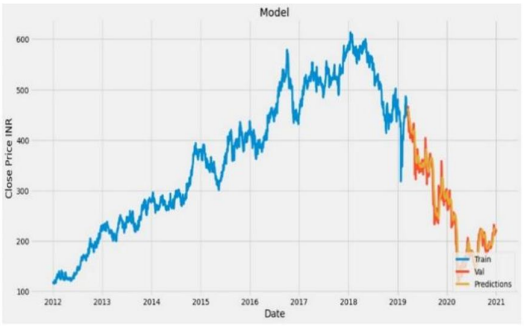

Figure 5. ZEE Ltd Stock Prediction

Figure 6. Hindustan Unilever Stock Prediction

the data and develop different algorithms that may help in finding appropriate stocks values. Predicting the stock market was a time-consuming and laborious procedure a few years or even a decade ago. However, with the application of machine learning for stock market forecasts, the procedure has become much simpler. Machine learning not only saves time and resources but also outperforms people in terms of performance. it will always prefer to use a trained computer algorithm since it will advise values based only on facts, numbers, and data and will not factor in emotions or prejudice.[3].

REFERENCES

[1] A Survey of Machine Learning Techniques Used on Indian Stock Market by Archit A. Vohra and Dr. Paresh J. Tanna 2020.

[2] Suganya R, and Sasi Kumar A 2020 Stock price prediction using tech news based soft computing approach International Journal of Advanced Trends in Computer Science and Engineering 9(2) 2049–2054.

TABLE I. PREDICTION OF STOCKS ON 27 APRIL 2022

STOCKS	PREDICTION (in RS)	CURRENT (in RS)	DIFFERENCE(%)
RELIANCE	1964.5118	1987.5	1.16%
TCS	2883.4312	2928.25	1.53%
HDFCBANK	1388.7642	1436.300049	3.31%
INFY	1190.0713	1255.800049	5.23%
HINDUNILVR	2315.3362	2395.399902	3.34%
ICICIBANK	532.36115	535.049988	0.50%
HDFC	1204.748	2558.649902	52.91%
BAJFINANCE	5376.618	5295.200195	1.54%
SBIN	277.55396	274.950012	0.95%
WIPRO	378.06885	386.25	2.12%
BHARTIARTL	587.26306	509.700012	15.22%
HCLTECH	841.9	946.15	11.02%
KOTAKBANK	1834.81	1995.59	8.06%
ASIANPAINT	2568.92	2764.5	7.07%
DMART	2708.1	2763.6	2.01%
ITC	207.37	209	0.78%
LTI	3020.116	3659.5	17.47%
BAJAJFINSV	7610.5	8906.39	14.55%
MARUTI	7674.64	7649.6	0.33%
TITAN	1510.41	1567.15	3.62%
ULTRACEMCO	5277.93	5288.14	0.19%
AXISBANK	636.7	620.45	2.62%
ADANIGREEN	1324.33	1052.59	25.82%
SUNPHARMA	598.08	592.34	0.97%
ADANITRANS	414.06	437.5	5.36%

STOCKS	PREDICTION (in RS)	CURRENT (in RS)	DIFFERENCE(%)
ADANIENT	460.08	479.54	4.06%
ONGC	97.72	93.05	5.02%
TECHM	975.41	973.2	0.23%
TATAMOTORS	191.58	183.85	4.20%
JSWSTEEL	379.82	387.2	1.91%
ADANIPORTS	482.35	483.75	0.29%
POWERGRID	144.03	142.38	1.16%
TATASTEEL	653.151	643.65	1.48%
HINDZINC	236.61	239.05	1.02%
HDFCLIFE	657.984	676.5	2.74%
LTI	3101.01	3659.5	15.26%
VEDL	157.79	161.39	2.23%
PIDILITIND	1636.46	1765.65	7.32%
DIVISLAB	3745.31	3841.89	2.51%
NTPC	101.93	99.34	2.61%
SBILIFE	911.62	904.25	0.82%
ZOMATO	142.64	137.39	3.82%
HINDALCO	236.3	240.55	1.77%
GRASIM	918.72	927.84	0.98%
IOC	88.24	90.94	2.97%
M&M	721.25	720.59	0.09%
DABUR	504.75	534	5.48%

[3] Srinivas K, Srinivasa Rao V, and Sreemalli M 2018 Binarized genetic algorithm with neural network for stock market prediction ACM International Conference Proceeding Series 8–11.

[4] Nayak A, Pai M M M, and Pai R M 2016 Prediction Models for Indian Stock Market Procedia Computer Science 89 441–449.

[5] Nair B B, and Mohandas V P 2015 An intelligent recommender system for stock trading Intelligent Decision Technologies 9(3) 243–269

[6] Nayak R K, Mishra D and Rath A K 2015 A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices Applied Soft Computing Journal 35 670–680.