Grenze International Journal of Engineering and Technology, June Issue

Algorithmic Trading using Machine Learning

Satvik Shayank[1], Sujit Saroj[2], Aditya Singh[3] and Dr. Vijaya Pinjarkar[4]
[1-4]Department of Information Technology, K J Somaiya Institute of Technology, Mumbai – 400022

***Abstract*—Predicting the value of a company's stocks is the motive of stock market predictions. Today's market use Machine Learning to forecast current market values by training on their prior values. As part of the machine learning process, various models are used to facilitate and authenticate forecasting. In the paper, we present results from machine learning and regression based on LSTMs for predicting stock value. A number of factors come into play, such as the open, the close, the low, the high, and the volume. Data quality, overfitting, and interpretability are examined as challenges and limitations of machine learning in algorithmic trading. Our final section discusses the future directions for algorithmic trading and machine learning research. Fusion can be viewed as a method of combining data or qualities in general in order to enhance prediction using a combinational method that can support one another. Financial market machine learning models are also examined for interpretability. The final part of our discussion discusses the future of algorithmic trading and machine learning research. Fusion can be viewed as a method of combining data or qualities in general in order to enhance prediction using a combinational method that can support one another. Ultimately, the integration of machine learning into algorithmic trading has the potential to revolutionize financial markets and improve trading decision-making efficiency.**

I. INTRODUCTION

One of the most well-liked investments is the stock market because of its huge profit margins. People looked for techniques and tools which maximise earning and minimising risk the volume of trading and investing rose. The two markets are National Stock Exchange and the Bombay Stock Exchange, where majority of trading on Indian Stock Market takes place. These systems can operate at high speeds and can process large amounts of data, making them popular among institutional investors and hedge funds.

Machine learning techniques have become increasingly popular in algorithmic trading due to their ability to analyze large amounts of data, detect patterns, and make predictions. Machine learning algorithms can learn from historical data and using it we forecast future event. It will help to take a better decision. Data mining is a approach intended to analyse data to find important insights. The nature of the stock market is quite unpredictable. stock price changes that occur practically instantaneously. Financial experts who buy stocks may not be aware of all the elements, such as inflation and economic growth, that influence stock prices. They are confuse about which stocks to buy and sell. They are simple to manipulate for stock brokers. Next, the model is examined using the test data. For this hypothesis, both LSTM and regression models are used individually. Regression includes reducing error,

and LSTM helps with long-term memory retention of the data and findings.

The graphs of price fluctuation over time (for a model based on regression) and in the middle of actual and projected price (for a model based on LSTM) are then displayed. Given its rarity, the temporal characteristic of stock data is regarded as an essential component of the prediction model. This is thought to be the primary aspect

Grenze ID: 01.GIJET.9.2.78

to be assessed using prophet, an open source library. A number of time series forecasting systems utilise this method, which was created by Facebook team. When there is a probability for seasonal results, it is frequently used. The Prophet was initially developed to generate accurate sales prediction. There are several strategies for forecasting time series. You can use MAPE to generate exact forecasts. Facebook Prophet make compare in middle of various activity and time activities. The exponential smoothness employs the similar tactic of modelling the season as an add-on. A multivariate analysis model called Logistic regression which used to forecast occurrence using forecast variable. LR applications have been widely employed in the fields of investing, corporate finance, and banking. A lot of scholars used multivariate discriminant analysis for the default-prediction model (MDA). They suggested a brand-new genetic algorithm-based binary classification technique for predicting corporate failure and suggested validating its predictive accuracy through empirical research.

II. Related Work

In a literature review, machine learning is extensively in Stock Market forecasting around the world. Modern prediction methodologies fail to compare favorably with machine learning approaches when it comes to speed and accuracy when compared to machine learning. Several studies have been conducted worldwide in this area.
A research conduct by S.H Adil, K. Raza, M.Z. Usmani and S. A. Ali [1] and K.Raza [2] presented their findings on the machine learning application of latest developments provides evidence to support these algorithms.
By utilizing deep neural networks, H. Gunduz, Z. Cataaltepe [3] , and Y. Yaslan[4] forecasted stock price. Similar to this, S. Waheed, M. Billah[5] and Hanifa A. proposed more enhancements to stock forecasting using Neural Networks by employing a practice method that they independently created. S.M Sundaram and K.V Sujatha[6] suggested enlightening methods for allocating with abnormal circumstances that frequently follow through system operation and might disrupt operations or result in indecent predicts.
Iu,G Lia and Y. Ding similar study was done [7] 2018 Security in cyberspace: first international conference, and create a model that uses LSTM to predict stocks that can be improved for accuracy in the future. S. Mohammed, E. M. El-Alfy and K.A. Althelya[9] conducted additional studies to examine the feasibility of using deep learning algorithms for stock forecasting. The method of a systematic literature review framework by Parris and Peachey (2013) is utilised to address the study questions. First, information pertinent to the inquiry was gathered. Second, this data was analysed and assessed. Anirudh Ragavender. R.M, D. Divya's ,Lilly Sheeba . S, Neha Gupta [10] method emphasizes the importance of time in improving prediction accuracy, by utilizing the prophet library.
In order to visualize how time has affected stock prices, we constructed a variety of graphs based on the dataset. Research done by Ramaswamy Seethalakshmi[11] in which Technology development enhances research into the stock and share market sector. Decision-making is improved by variety of statistics and machine learning method. There has been a lot of research done on the feature of stock price prediction based on past prices and volume. In this study, nifty50 Index performance measures are assessed using statistical techniques in a R environment. The study's findings outperform those of earlier approaches. Linear regression is the foundation of traditional financial market analysis techniques. Stock market closing prices can be predicted using the finest single variables are the topic of this essay. The reason of this report to study is to categorize the precise variables affecting greatest influence on closing value prediction. By Subba Rao Polamuri, K. Srinivas, A. Krishna Mohan [12] In this research, we tried to undertake novel work by using a machine learning methodology to sense or anticipate the sensex stock market's behaviour tracking. A variety of machine learning methods, support vector , including linear regression, extra tree regressors ,decision trees, and random forest regressors, are effectively used to forecast stock prices and provide an overview of securities trading . As a result of the closing price and stock price, we estimated the stock's price . An algorithm with high accuracy is one that performs a process of comparison for the efficiency of each model and is ultimately regarded as the best algorithm for stock price prediction. [13] In this paper it prompts investigation into the most efficient prediction model that yields the most precise forecast with the lowest proportion of mistake. Time series data are used to represent stock prices, and neural networks are trained to find patterns from trends in the historical data. To increase the accuracy of stock price prediction, this system used an LSTM algorithm. Research was done by Mohammed Ali[14] this paper will go over how to use the prophet and LSTM models for stock price implementation, forecasting, and research. This task and procedure are extremely difficult and unclear. . Although it is impossible to forcast stock prices due to their ambiguous nature, this research seeks to use forecasting and data analysis to do so.

The prophet algorithm explains three per parameters mainly seasonality, holidays and trends. Every parameter illuminate the significance held by period in financial market forcasting.
The Santander Group stock prices from the Euronext Stock Exchange are included in the dataset that is being utilised for the prophey. It includes four years' worth of time series data as well as information about the stock's characteristics each day, including its peak, lowest, open, and final values. The prophet algorithm, which is designed using the training set give input, is also used by the data pipeline. To predict the future value of a stock price, machine learning techniques like LSTM are currently used in prediction of stock price . Machine-learning algorithms are aimed at providing the most accurate stock price predictions. In this proposed system, LSTM detects the changes in the stock price's behavior over a specified time period. Offer a normalization based on machine learning for predicting stock prices. Yahoo Finance provided the dataset that was used for the analysis. It has roughly 11 lakh records with the necessary Stock price and other pertinent information. The information showed the stock price for each day of the year at various intervals. It includes a variety of information, including date, symbol, high price, low price, close price, open price and volume. In this case, data for just single company was taken into account. The data was accessible through a CSV file, In Python Pandas, a frame was created by reading the file and converting it. Stock data divided into training and testing sets after being normalised using the Python sklearn module. 20% of the available dataset was kept for the experiment set.

III. METHODOLOGY

Following diagram is system architecture of our model:

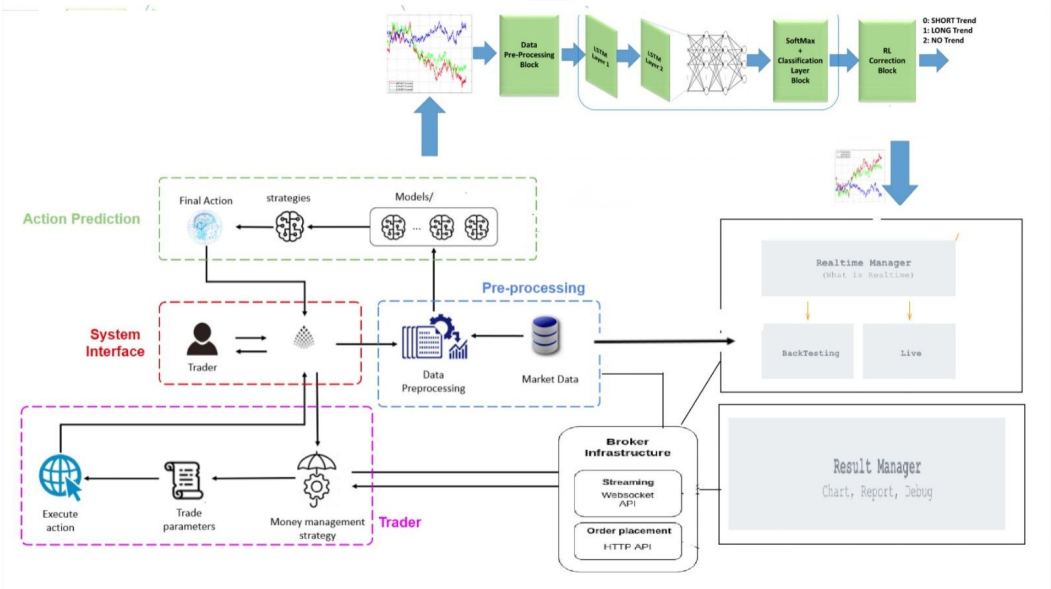


Fig. 1 System Architecture

The system presented here composes of five modules: -

- Input as Dataset
- Pre processing
- Data splitting
- Build & Model train Lstm, CNN and Prophet
- Predicted result (Output)

Data attributes like price of open, high, low, close, adjusted close price are used to train models by applying preprocessing techniques such as normalization and one-hot encoding. Stock data is split in parts, training and testing.
Opening price: Starting price refers to the price in the case of the Nifty 50 stock exchange, the price of each stock when trading begins. In order to predict the stock's movement on a daily basis, the opening price is a good indicator.
There is no need for the beginning price to match the previous day's closing price since the stock market is like an auction where bid and ask meet to make transactions with bidding their highest price.

Highest and Lowest price: Based on the previous day's highs and lows, we can get an idea of how the shares typically fluctuate throughout the day and how this will ultimately affect the closing price. Furthermore, it shows how each share's cyclical movement has changed over time.
Closing price: An individual stock's closing price on a specific trading day can be calculated in the following ways , any distributions made prior to the next day's open are taken into account. Often, adjusted closing prices are used to analyze historical returns in detail.

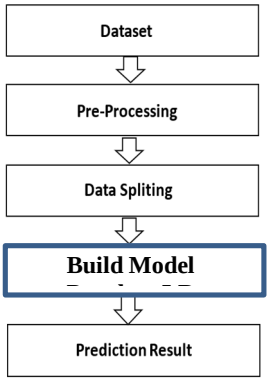


Fig 2 Proposed workflow

A. Dataset

Taken from Yahoo Finance stock data from Kaggle.com.It's attribute is price of open, low, high, close, adjest closed etc. After taking data we perform preprocessing and then split the data in ration of 80:20.

B. Data pre-processing

In order for the machine to readily parse the data in our dataset, we will convert it in this stage so that it can be understood by the machine. The data would be divided into a training and testing dataset.

C. Algorithms Used

LSTM :
It is used to find out recall insights after a period of time. A memory cell's updated state and its newly introduced input can be used to compute its new output when another input is introduced. Neuron output signals are sent from each processing component to as many output connections as required. In the neuron output signal, any mathematical type may be represented. Gradient descent is no longer present in Recurrent Neural Networks, so they cannot learn from historical data. LSTMs, also known as long-term memory networks, have been used to tackle this problem. Long Short term memory was first proposed by Hoch Reiter and Schmid Huber in 1997 as a solution to these problems. The output is decided by a combination of the cell state (representing the prior information) and every time an input is received, the cell state is updated. A new input can be used to compute a new output when a new input is introduced into the memory cell. A high-performance computing cluster was used to implement the algorithm using Theano and Keras. An algorithm called "fit" specifies how the LSTM model is created. In addition, it takes into account the amount of epochs and fits to a given dataset, no of neurons and no of blocks in the model. In order to construct a network, the mathematical notation of Theano must be followed. Building a model requires defining both the loss function and the optimization procedure jointly. The LSTM model is used to forecast the following step in the dataset by developing a compact function. First, an LSTM model is constructed using a predetermined training dataset, epochs, and neurons in the algorithm's active section. Forecasts are also being made for the training data.
Linear Regression :
In linear regression, two variables are analyse to find out single relationship. A trader can use linear regression to locate important value point, which can be entry, stop or exit value. Machine learning is process of teaching a compute to recognise a data. It's general equation: $y = w1 * x+b$
Linear regression algorithm steps:
Step-1: the algorithm is data collection.
Step-2: Define the alpha value of the hyperparameter.
Step-3: perform gradient function to define values of b.
Step4: Use formula $y = w1*x + b$ to define the value of y.

Support Vector Regression:
To forecast discrete values, support vector regression is used as an approach for supervised learning. It is the same theory that governs SVMs and Support Vector Regression. SVR's fundamental principle is finding best fit line. Similar to other regression technique it match the value in bound and oppose to reducing variation between projected and real value. Additionally, Support Vector Machines can be used as regression techniques, which can be divided based on their unique properties (maximum margin). There are only a few difference between SVR and SVM for classification. It provides an endless number of possible outcomes, making it extremely challenging to obtain information on the tips of one's fingers. Also, this regression contains a tolerance margin for the "epsilon" of the SVM, which is already defined during the hypothesis process.
Using SVR, closing stock prices are predicted in this study. Using zero average returns on stocks and a given variance, the results are compared with those obtained by the random walk model. A model's price predictions are evaluated according to their Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). RMSE and MAPE can be calculated as follows:

$$\mathrm{MAPE} = \frac{1}{T}\sum_{i=1}^{T}\left|\frac{d_i - \hat{d}_i}{d_i}\right| \quad \text{Eq. (1).}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}\left(d_i - \hat{d}_i\right)^2} \quad \text{Eq. (2).}$$

There are many types of indicators to choose from, but the simplest are moving averages, simple to calculate and understand. Moving Averages (SMAs) are calculated by taking the arithmetic mean of previous prices, according to

$$\mathrm{SMA} = \frac{1}{T}\sum_{i=1}^{T} Cl_i \quad \text{Eq. (3).}$$

Using SVMs, you can classify observations between groups by mapping their independent variables into an array of dimensions. The model uses non-linear classification thresholds to construct a linear model, mapping variables over a broader range of dimensions. An optimal hyperplane is used to divide classes based on N observations, x representing the independent variable vector, and y representing the classification of each sample. Therefore, the classification hyperplane can be calculated using Eq. (4),

$$\mathbf{w}^T \phi\left(\mathbf{x}_k\right) + b = 0 \quad \text{Eq. (4),}$$

To predict asset prices, SVR is used to get a regression model. Here are the notations used in the SVR model:

- N: total samples; • T: total test samples; • d: real sample value; • P: period (minutes or days);
- $\hat{d}$: value estimated by the model; • Cl: closing price for the period; • r: return.

Stock	Price	Prediction price	Confidence score
Adobe inc.	436.37	425.22	0.26
Amazon inc.	124.25	118.30	0.35
Alphabet Inc.	125.3	121.56	0.47
Visa Inc.	228.79	220.45	0.42
Walmart Inc.	148.2	141.54	0.38

The table presents the current prices of various stocks, along with their corresponding predicted prices and confidence scores
Prophet:
It is method to predict time series data it is accountable for non-linear trend which can be yearly, monthly, day. Prophets is good to use historical data with various seasons and time series with high seasonality. A stock prediction dataset can be used to implement a variety of time series approaches, however most of these techniques demand extensive pre-processing of the data before building the model. An open-source library Prophet was created for time-series dataset forecasting. Time series forecasting models can be created using the prophet algorithm. In contrast to conventional techniques, it tries to fit additive regression models. In addition, the algorithm can receive data in a variety of formats. Prophet (value to forecast) accepts only data frames with a date stamp (ds) and a year (y). A new attribute "ds," "y," was created to store the dates and values. For timestamps, ds (date stamp) should be in a Pandas format; for dates, ds (date stamp) should be YYYY-MM-DD HH:MM:SS. Forecasts must be represented numerically in the y column.The analysis of time series can be

performed using a variety of time series techniques the majority of these techniques do not require a stock prediction dataset require much pre-processing of the data before constructing the model.
A prophet library is an open-source software that serves the purpose of prediction for time-series datasets. It is designed to be easy to use the hyperparameters of the model are automatically found. A model for making accurate forecasts of trends in data structure is based on the seasonal cycle.
The prophet is an additive model with the following. components: $y(t) = g(t) + s(t) + h(t) + \varepsilon_t$
Forecasting is done using the prophet algorithm. An analysis of time series data using models. The method differs from the traditional method, By fitting additive regression models, it tries to predict the future. It also offers great flexibility in terms of the data it accepts algorithms are fed with this information.

trend	yhat_lower	yhat_upper	trend_lower	trend_upper	additive_terms	additive_terms_lower	additive_terms_upper	weekly	weekly_lower	weekly_upper	yearly	yearly_lower	yearly_upper
106.0346	-68.9947	272.9836	-67.9015	269.6713	1.5420	1.5420	1.5420	0.0608	0.0608	0.0608	1.4812	1.4812	1.4812
105.4728	-69.8590	276.5036	-69.1182	269.6606	1.9090	1.9090	1.9090	0.4488	0.4488	0.4488	1.4603	1.4603	1.4603
104.9109	-73.9427	278.0574	-70.3348	269.6499	1.0185	1.0185	1.0185	-0.3437	-0.3437	-0.3437	1.3621	1.3621	1.3621
104.3491	-73.5496	279.0097	-71.5514	269.6391	0.3757	0.3757	0.3757	-0.8104	-0.8104	-0.8104	1.1861	1.1861	1.1861
103.7872	-77.6565	276.8626	-72.7681	269.6284	-2.3128	-2.3128	-2.3128	-3.2454	-3.2454	-3.2454	0.9326	0.9326	0.9326

Fig.3 Prophet algorithm trend of stock performance

Forecast values " yhat " and actual values "y" are in the data frame, and the cut-off date is specified in the performance_ metrics() function.

D. Backtesting

The backtesting process involves simulating how a strategy would have performed in the past based on previous market data to evaluate the performance of an algorithmic trading strategy. Prior to applying trading algorithms and strategies to live markets, this process is crucial for developing and refining them.
Backtesting involves feeding historical market data into an algorithm that generates simulated trades based on the rules and parameters defined by the developer. Simulated trade results are then compared to actual market data to determine how well the algorithm would have performed if it had been used in real-time.
In conducting a backtest, however, several factors must be considered, such as the quality and completeness of historical data, the accuracy of the trading rules and parameters used, and trading costs and slippage that would have occurred. It is important to carefully consider these factors when conducting a backtest to ensure that the results are accurate.
In spite of the fact that backtesting can provide valuable insights into a trading strategy's performance, it does not guarantee future results. Performance is not necessarily indicative of future performance, as market conditions can change rapidly and unexpectedly. Based on real-time market data and feedback, it is essential to monitor and refine the trading algorithm.
In spite of the fact that backtesting can provide valuable insights into a trading strategy's performance, it does not guarantee future results. Performance is not necessarily indicative of future performance, as market conditions can change rapidly and unexpectedly. Based on real-time market data and feedback, it is essential to monitor and refine the trading algorithm.

E. Result

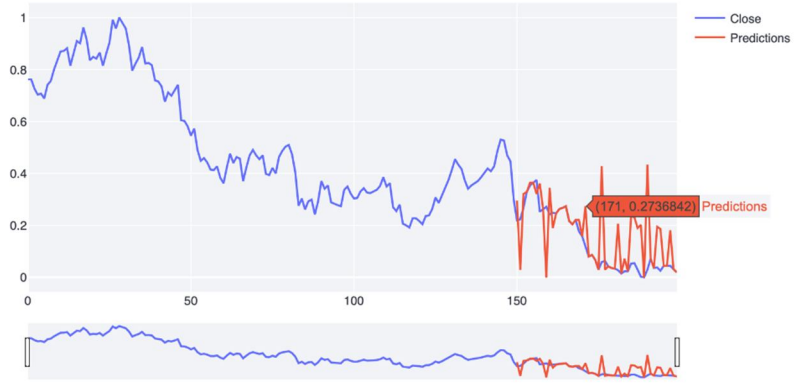


Fig.5 Linear Regression model

In the above figure 5, we have defined the stock price in the limit from 0 to 1 for 200 days. The red line represents stock prediction for last 50 days. It is difficult to predict stock prices for the last 20 days due to volatility.

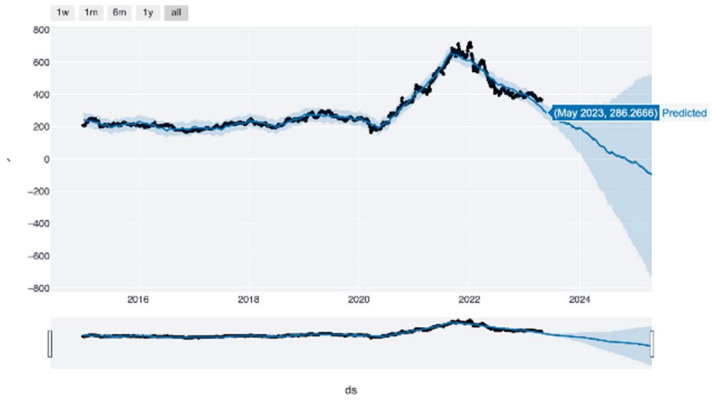


Fig 6 Forecasting prophet

In the the above figure 6, we had data till mid 2022, and after that the data generated is by the prophet algorithm. The prediction here is shown on the basis of seasonal stock price movements. As of May, 2023, the predicted value of the stock is Rs.286.2666

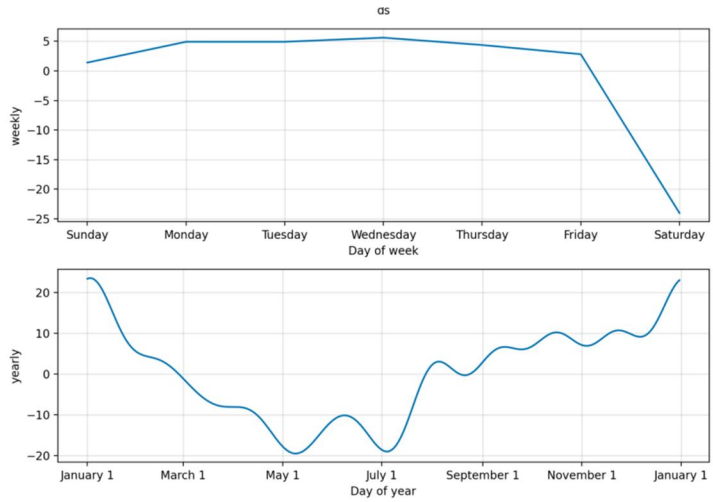


Fig.7 Forecast Component using Prophet

According to the model in figure 7, the trend of the curve is almost upward, and that has been the case for Bharti Airtel stock. Bharti Airtel has been a profitable company, and hence stock prices rise up.

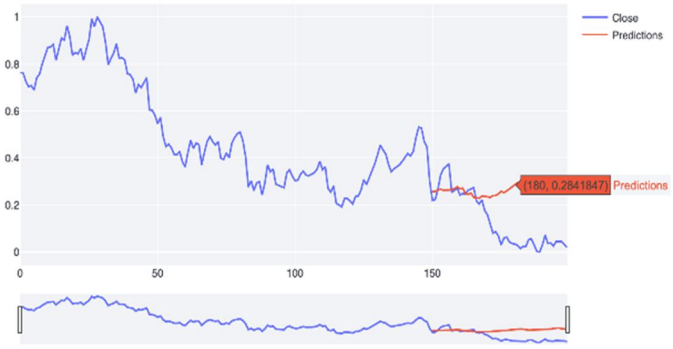
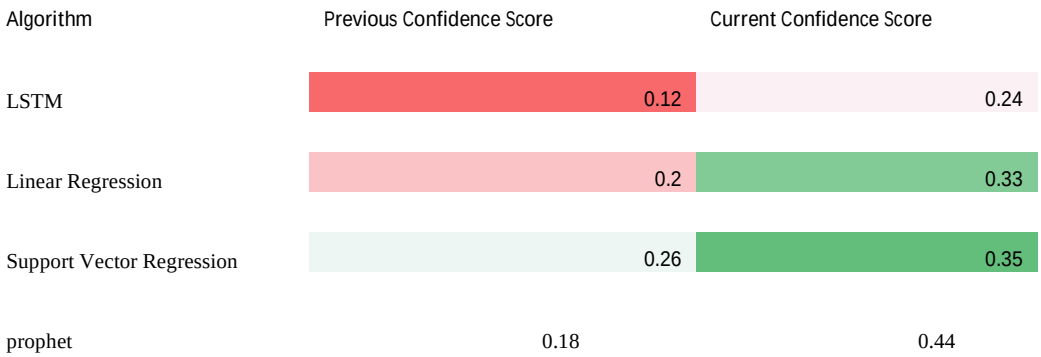


Fig.8 LSTM model

In the above figure 8, we have defined the stock price in the limit from 0 to 1 for 200 days. The red line represents stock prediction for last 50 days. It is difficult to predict stock prices for the last 20 days due to volatility. The predicted stock price for 180th day is Rs.0.284.

IV. OBSERVATION



From the above table we can phase out the following comparisons of our model from the existing models:

- For LSTM algorithm, the confidence score compared to the existing systems has increased from 0.12 to 0.24, with a percentage increase of 50%.
- For Linear Regression algorithm, the confidence score compared to the existing systems has increased from 0.2 to 0.33, with a percentage increase of 39.4%.
- For Support Vector Regression algorithm, the confidence score compared to the existing systems has increased from 0.26 to 0.35, with a percentage increase of 25.7%.
- For Prophet algorithm, the confidence score compared to the existing systems has increased from 0.18 to 0.44, with a percentage increase of 59.1%.

V. CONCLUSION AND FUTURE SCOPE

Stock brokers and investors are assisted in making financial investments on the stock market by our research study. In a dynamic market like the stock market, prediction plays an important role, which is a complex and challenging process. The aim of project to summarize important ML method relevant to the prediction stock prices . In the study, SVMs are suggested as a way to generate accurate market predictions and stock analysis, while the publication suggests linear regression and logistic regression. There is a limitation to this result due to the need for datasets to be classification-friendly for use in prediction. In addition, we have employed five algorithms that aid in predicting the stock price and confidence score.

Using stock price data, we tested a back propagation neural network approach using gradient descent optimization. The gradient descent approach can be used to obtain the values of the training cycle and learning rate by utilizing the model-based neural network algorithm. Best predictions are those whose computational results are the smallest. To predict the asset price, SVR is used to get regression model and forecasting is done by the Prophet algorithm.

By analyzing all the historical data that we have archived, we could find trends, what worked and what didn't work in the entire history of trading. Furthermore, it could learn to predict future markets effortlessly while trading multiple accounts and strategies to spread risk and reject or accept real-time bids and offers. Trading strategy rules can be adjusted to match market conditions if the market does not favor your trading strategy. In addition to saving time and eliminating any possibility of a slight gap in time or error, it could be capable of checking out multiple market conditions at once across the globe. If markets become intelligent, markets may not crash if there's a mistake in buying or selling, or market volatility may dip and the exchange will rise to the challenge without human intervention. We can thus expect the elimination of certain discrepancies like Flash Crashes. In order to increase efficiency and facilitate communication, algorithms could be programmed directly into the chips The programming of specific global regulations and kill-switches could also be performed. It would be incredible if there was a system this powerful, with so much reach and connectivity, with connectivity like the Internet of Things, and speed of internet and processing power, integrating structured and unstructured

data and integrating structured and unstructured data and using one algorithmic engine to combine real-time global news feeds, social media and current and historical stock data across the globe. This can also be used to make portfolios.

REFERENCES

[1] Saxena S, Agarwal N, Chouhan L, Stock prediction using machine learning, international conference on secure cyber computing, IEEE, 2018
[2] Soni V, Pahwa N, Stock Market prediction using machine learning, International Journal publication, 2017
[3] Singh U, Bhuriya D, Sharma A, An analysis of Stock prediction using Machine Learning, International conference of Electronics , Communication and Aerospace, 20 Apr 2017s
[4] Chaudari K, Thakar A, Fusion in Stock Market Prediction, The fusion of Information, Jan 2021
[5] Raza K, Adil S, Usman M, A Machine Learning approach to Stock Market Prediction, 3rd conference, (ICCINS), 15 Aug 2016
[6] Raza K , Predicting stock market behavior using Machine Learning techniques, The international conference on computer and information science (ICCINS), 15 Aug 2016.
[7] Cataltep Z, Gundu H, Prediction of Stock market direction using Deep Neural Network, IEEE, 15 May 2017
[8] Boluk S, Gudelak M, Detecting trend in 2-D CNNs through Depp Learning, IEEE (SSCI), 2017
[9] Riordan R, Hendershott T, Algorithmic Trading and Information, Califonia, 2009
[10] Neha Gupta, Dr. lilly Sheeba, , Time Series Model for Stock Market Prediction using Prophet, Chennai, 2021
[11] Ramasway Seetalakshmi, Analysis of Stock Market predictior varible using LR , Jan 2018
[12] K Srinivs, Subba Rao Sock prediction using Random forest, Sep 2019
[13] Dr. Miralpatel, Drashti Talati, Stock prediction using LSTM, Vallabh Vidhyanagar , Jun 2022
[14] Mohammd ali alshara , Prophet vs LSTM model for stock forcasting, Islamic university Riyadh, Saudi Arabia, Feb 2022
[15] Zheng Z, i Y, algorithmic Trading base on Reinforcement learning , IEEE, 2019
[16] Vinayakumar R, Selvin S, Comarison of LSTM, CNN- slidin and RNN models for stock price prediction, ICACCI, 217
[17] A. V. Devadoss, T. A. A Ligori , Multilayer perception forcasting of stock price, Int J computer Algo.
[18] R. J. Hyndman, J.G De Gooijer , Time series furcating , International Journal for forecasting, vol 22