International Journal of Scientific Research and Engineering Development-– Volume 6 Issue 2, Mar-Apr 2023

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

ALGORITHMIC STOCK MARKET PREDICTION

Mhaske . V. D.*,Pratik Karale**, Rushikesh Shinde**,Pritee Khaire**,Rutuja Dhandore**

*(Assistant Professor, Computer Engineering , SVPM's College of Engineering Malegaon(bk), Baramati
Email: vdmhaske@engg.svpm.org.in)

**(UG Students,Computer Engineering , SVPM's College of Engineering Malegaon(bk), Baramati
Email:pratikkarale692002@gmail.com ,rushi6270@gmail.com,priteekhaire88@gmail.com,rutujadhandore7@gmail.com)

----------**********----------

Abstract:

The stock market is one of the best way to earn money and invest. An organization's stock price is a measure not only of its current stock value, but also of its future performance. The stock market is where buyers and sellers trade equity stocks. The company is a fully licensed organization, and its stock trading members are registered with SEBI. It is difficult to predict or predict future stock price values.In previous paper we observed out that strategies like random forest, SVM and CNN algorithms are used but instead of that now we are using the LSTM model which gives more accurate results than those algorithms. This article focuses on comparing the use and effectiveness of various stock market forecasting models, including long-term memory, neural network techniques, and reinforcement learning techniques. Reinforcement learning and LSTMs achieve accurate results in the stock market, according to the results of this research paper.

Keywords **—LSTM(Long short-term memory) , SEBI(Securities and ExchangeBoard of India),SVM (Support Vector Machine).**

----------**********----------

I. INTRODUCTION

Predictions of stock prices are very important to a lot of corporation related people and the general public. A career in the stock market can make or break someone's financial life. Using historical data, algorithmic predictions and models can be used to predict the future. Predicting the future has been a daunting endeavor that many people have struggled to comprehend. When it involves money and risks, like speculating on the stock market, this kind of prediction is even more appealing. Stock market forecasts are the subject of research from a wide range of disciplines, including business and computer science. Different strategies, algorithms, and combinations of indicators have all been tried by researchers to predict the market. A prediction model's attribute is based on factors that can affect market performance. The Long ShortTerm Memory (LSTM) structure is one of many Recurrent Neural Network (RNN) structures. In the hidden network layer, LSTM transforms the most useful memory cells into the conventional artificial neurons. Networks can better associate memory with remote input over time thanks to these memory cells, so it's important to understand how strong data with great predictive power form over time. News articles, Twitter data, Google data, and Wikipedia data are just a few of the many data sources that have been used in daily research on stock price forecasts. Stock prices have changed as a result of these external factors, as well as stock prices and technology indicators. In today's society, it is unknown how stock prices can be made more accurate. A time series is a arrangement of data from a variety of disciplines, including economics, finance, social science, engineering, and physics. economics, engineering, finance, and physics. Predicting price trends is extremely challenging when dealing with complexity of this kind.

Predicting a series of time is primarily used to build future value counterfeit models based on their previous values. This amounts to revealing the distribution of conditional event as a function of foresight since the relationship between past and future recognition is frequently unclear.

There are two different stock types. You may knowintraday trading through the phrase "day trading". Intraday traders frequently hold security position for multiple days up to weeks or months, but at least from one day to the next. LSTMs are mostly effective in solving sequence prediction issues, because they have the capacity to store historical data. This is important in our circumstances since a stock's historical price plays a key role in defining its future price. Although predicting a stock's actual price is hard, we can form a model that will predict whether it will increase or decrease. This project's primary goal is to forecast stock values utilising lstm.

All the concepts related to project is discussed in following section which includes Literature Survey where study of existing systems was done.In next section Proposed work and LSTM overview is discussed .After that detailed architecture followed by result analysis is explained.

II. LITERATURESURVEY

The evaluation of various systems and models to ascertain whether stock predictions could be finished based on actual stock prices was the primary focus of our literature review[5]. However, since we have not been able to notice any possibility of a modification in this stock price prediction, we have made the decision to examine the plans that are in place, evaluate the major issues, and make improvements to ourselves. We found LSTM after doing a quick search for common solutions to the above problem. Time sequence data from stock market prices of the stock and connected fields variables over a number of years is collected afterthe decision has been made to use the neural network to make stock calculations[8][9].

Forecasting stock prices using morphological similarity clustering and -level temporal memory. A stock mining clustering method that associations morphological similarity distance (MSD) and kmeans clustering. In terms of short-term calculation, the C-HTM performance is effective than all baseline models but for long-term prediction, the performance of C-HTM is very deprived than all basicmodels [3]. Inventory prediction based on convolutional neural network and feature selection Closed Bidirectional recurrent unit display that the fusion model (FS-CNN-BGRU) which works better than other simple models have a certain reference value. Bidirectional Gate Recurrent Unit (BGRU) can procedure data with time series, so it can have improved performance for data with time series characteristics, but in experiments, they used a hybrid model of CNN, CNN-GRU and FS-CNN-BGRU, which makes it complicated and difficult to understand[1].

Predicting Stock Prices using Limit Order Books Using Convolutional Neural Networks, This work proposes a deep learning tacticgrounded on Convolutional Neural Networks (CNN) to predict stock price movements using as an input to book no. large scale and high frequency. Alike multilayer neural networks and support vector machines, it shows that CNNs are better suited for this type of task but it requires large training data and does not encode object positions and orientations[6]. Stock price prediction using LSTM, RNN, CNN with Sliding Window predicts the price of NSE listed companies using three different deep learning architectures, equivalences their performance, and smears them. A sliding window method for forecasting future value in the short term. These methods are used to focus on data from multiple companies, predict prices using an internal model and compareperformance [4].

SVM with adaptive parameters can attain higher generalization performance in financial forecasts, and it's better than the standard SVM useless support vectors. It is also proficient of handling high-dimensional data, which evidences to be a great help seeing its use and claim in the field of machine learning, but SVM algorithm is not

suitable for large datasets. If the dataset has more noise, SVM will not perform well[7].

TABLE I
COMPARISONS BETWEEN MODELS

Model /Parameter	CNN	LSTM	RNN	GRU's	SVM
Long term Prediction	poor	Better than RNN and GRU's	Limited	better	good
Gates	None	Input,Output , forget	None	Update ,reset	none
Hidden State	Multiple	Multiple	Single	Single	Single
Structure	Simple	More Complex	Simple	Simple	Complex

III. PROPOSED SYSTEM

We seek to determine the next day's closing value's precise value in the proposed system so that investors can buy or sell shares with confidence. An artificial neural network used in deep learning is called Long Short Term Memory (LSTM). An advanced neural network called the LSTM has a memory cell that temporarily retains a tiny amount of data for future use.

We propose making use of the LSTM (Long Short Term Memory) algorithm to effectively predict stock prices.

1. LSTM overview

A type of recurrent neural network (RNN) is the LSTM. RNNs are a powerful kind of artificial neural network that can keep memories of the input inside themselves. As a result, they are ideal for solving problems involving time series or sequential data. However, vanishing gradientis acommon issue with RNNs that can cause model learning to become excessively sluggish or even stop altogether. This issue was addressed in the 1990s by the development of LSTMs. LSTMs can learn from inputs that are separated by long time lags and have longer memories.

Three gates make up an LSTM: an input gate that decides whether or not to accept new input, a forget gate that clears out irrelevant data, and an output gate that decides what data to send out. Based on the sigmoid function, these three analog gates operate between 0 and 1. Figure shows these three sigmoid gates below. The state of the cell can be seen represented by a horizontal line that runs through it.

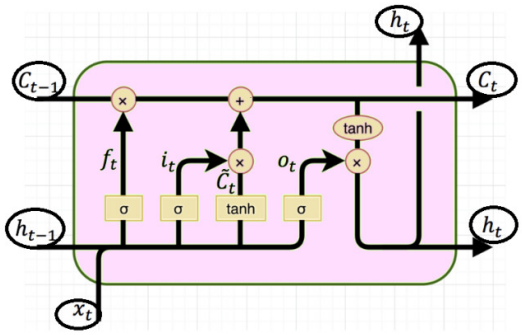

Fig.1 LSTM memory cell

2. Advantages of LSTM

An LSTM's main advantage is its ability to learn context-specific temporal dependence. Without explicitly using an activation function within the recurrent components, each LSTM unit remembers data for moreover a long or short period of time (hence the name).

It is important to observe that any cell state is increased only by the forget gate output, which ranges between 0 and 1. The forget gate in an LSTM cell is in responsibility of both the weights and the activation function of the cell state. As a result, information from a previous cell state can pass through a cell unchanged rather than increasing or decreasing exponentially at each time-stepor layer, and the weights can converge to their optimal values in a reasonable amount of time. Because the value stored in a memory cell is not iteratively modified, the gradient does not vanish when trained with backpropagation.

3. Working of LSTM

An important component of an LSTM model is a memory cell known as a "cell state" that maintains its state crossway time. The horizontal line that passes through the upper of the diagram below indicates the cell state. It can be visualized as a

conveyor belt across which data merely and unaltered passes. In LSTMs, gates regulate the adding and deletion of data from the cell state. Information may be able to arrive and leave the cell from these gates. The mechanism is assisted by a sigmoid neural network layer and a pointwise multiplication operation.Amongst zero and one, the sigmoid layer outputs a number, with zero representing "nothing should be let through" and one representing "all should be let through."

IV. OBJECTIVES

- To develop a prediction model for the businesses that are included in the NSE LIX15 index
- Using multilayer perceptron (MLP) Neural Network technique(LSTM).
- To compare the model with real data for its accuracy.
- Try to improve accuracy using Neural network.

Stock Price Prediction Using Machine Learning supports you in defining the future value of a company's stock and extra financial assets traded on an exchange. The goal of stock price prediction is to make a lot of money. Forecasting how the stock market will perform is a difficult task.

V. ARCHITECTURE

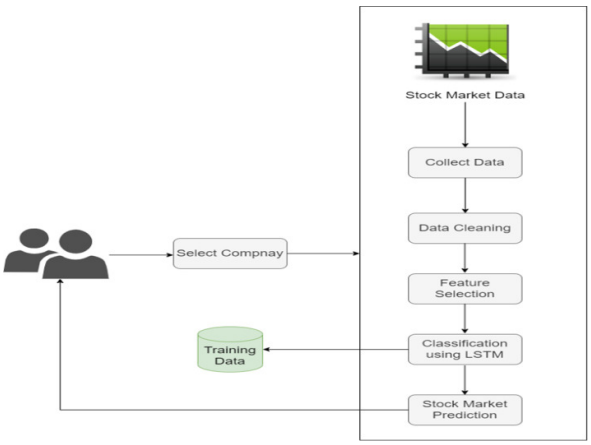

Fig. 1 Overall Architecture

STEP 1-COLLECT DATA AND DATA CLEANING

To begin, we will acquire the stock data by making use of yFinance. We can obtain free stock data from Yahoo Finance using the open-source Python library yFinance.

There were five features in the obtained data:

1. Date: Date of the stock's price.
2. Open: This is the stock's opening price at the start of each day's trading.
3. High: the amount at which the stock was traded at its highest through a given period (day).
4. Low: The lowest price at which the stock was traded over the course of a day.
5. Volume: what quantity of a specific financial asset has been operated over time.

The Pandas package and Python will be used to complete our data cleaning work. Dependent on the form of data and the goal you're trying to accomplish, data cleaning can vary greatly. In our situation, data validation is being done to ensure that the entered data is of the correct data type (numeric), and normalising missing and incomplete data to locate and complete any gaps in the dataset.

STEP 2 – FEATURE SELECTION

Investors and researchers always pay close attention to financial market analysis. The stock market trend is extremely intricate and influenced by a variety of factors. Therefore, it is crucial to determine the stock market's most significant factors. An example of such an algorithm is Feature Selection,which can eliminate duplicate and irrelevant features and identify the most important subset of features from which to construct the analytical model, so in our case volume and adj close features are removed in feature selection.

STEP 3 – STOCK PREDICTION USING LSTM

Data was then converted into a setupappropriate for usage with our forecasting model by carrying out the subsequent actions:

1. Converting time-series data into input-output elements for supervised learning
2. Grading the data to the [-1, +1] series.

The given data is divided into training and testing datasets; our LSTM model will be fitted on the train dataset, and the correctness of the fit will be assessed on the test dataset. One input layer with five neurons, 'n' hidden layers, and one output layer make up the structure of the LSTM network. When the model has been fitted to the training dataset, the validation set is used for hyper-parameter tweaking to determine the ideal values for parameters like the 'n' number of hidden layers.

The endofday price of any stock can be predicted by fitting the LSTM model to the training data.

There are two ways to execute this prediction:

1. Static: This straightforward but less precise method fits the model to all training data. Next, one by one, test results are used to anticipate each new time step.

2. Dynamic - a complicated, more precise method in which the model is updated for individual time step of the testing data as new remarks become available.

The RMSE (Root Mean Squared Error) metric can therefore be used to estimate the prediction model's accuracy in a reliable manner. This is because neural networks in general, including LSTM, have a propensity to provide inconsistent outputs when given identical data under various initial conditions.

We then carry out the model prediction numerous times (with various starting conditions), and we use the average RMSE as a measure of how well our configuration should perform on hypothetical real-world stock data. To put it another way, we will contrast our forecasts with the actual trends in stock price that can be deduced from past data.

VI. RESULT ANALYSIS

In below table different errors are calculated on different models which include RMSE (Root mean square error), MAE (mean absolute error), MSE (mean squared error).In RMSE lower the error value better the model , in MAE also lower the error rate better the model , in MSE lower the value better the model.

In case of SVM errors are higher than other models because it is not suitable for time series data. In Random Forest the difference between errors during training and testing shows inconsistency in model which is similar in GRU and LSTM+GRU also. In KNN errors during training and testing are too high because it is not capable of long term predictions. But in case of LSTM errors during training as well as testing are not that much different this is why LSTM shows better results in prediction.

TABLE 2
ERRORS ON PREDICTION

Algorithms	RMSE		MSE		MAE	
	Train	Test	Train	Test	Train	Test
SVM	37.27	36.34	1389.39	1320.79	31.28	27.23
Random Forest	15.94	36.39	254.20	1323.92	11.31	28.44
KNN	48.43	57.96	2345.29	3359.48	36.20	42.47
LSTM(200 epoch)	33.42	34.92	1118.76	1219.28	25.71	26.02
GRU(200 epoch)	24.43	48.30	597.04	2332.68	18.72	36.05
LSTM + GRU (200 epoch)	29.59	37.13	875.56	1378.43	21.97	27.93

Root means square error is mostly used to measures the quality of predictions. Itsusages Euclidean distance to show how distant predictions vary from measured correct values. The Mean Squared Error is a method that measures how close a regression line to a set of data points. The Mean Squared Error represents the expected value of the squared error loss as a hazard function. The typical, specifically the mean, of errors squared as of data which it relates to a function is use to calculate Mean Square Error. The amount of error in your measurements is expressed as Mean Absolute Error. It represents the transformation between the measured and actual values.

VII. EXPECTED RESULT

Our system will produce predictions of stock prices at the end of the day and indications on the basis of higher and lower threshold value whether he should Buy, Sell, or Hold a particular Stock.

VIII. CONCLUSIONS

It has led to the conclusion that the LSTM which is Deep learning technique can be used to predict the stock market with greater accuracy and efficiency. This system predicts market indices with increasing accuracy by training on their previous values, which are open, high, low, and close. In addition, the total market volume of the stock is provided.

This procedure is much simpler when machine learning is used for stock market forecasting. It saves time, resources, and outperforms people in terms of performance. In such unpredictable circumstances, institutions and share brokers will forecast the price in the future. This method of forecasting investment returns would be beneficial financially.

ACKNOWLEDGMENT

This paper could not have been written without the helpful advice and support of Prof. Mhaske. V. D. For this paper, we are grateful to Prof. Dr. Sinkar. Y. D., Head of Department, and Honourable Principal Prof. Dr. Mukane S.M. for their assistance.

REFERENCES

[1] Stock prediction based on bidirectional gated recurrent unit with convolutional neural network and feature selection, Qihang Zhou, Changjun Zhou, Xiao Wang,2022

[2] Stock Prediction Based on Genetic Algorithm Feature Selection and Long Short-Term Memory Neural Network d December 4, 2020 ,SHILE CHEN AND CHANGJUN ZHOU

[3] Stock Price Prediction Based on Morphological Similarity Clustering and Hierarchical Temporal Memory,Xingqi Wang, Kai Yang, Tailian Liu,2021

[4] Stock price prediction using LSTM, RNN and CNN-sliding window model Sreelekshmy Selvin, R. Vinayakumar, E. Gopalakrishnan, V. Menon, K. Soman,2017

[5] Nandakumar, R., Uttamraj, K. R., Vishal, R., & Lokeswari, Y. V. (2018). Stock price prediction using long short term memory. International Research Journal of Engineering and Technology, 5(03).

[6] Forecasting Stock Prices from the Limit Order Book Using Convolutional Neural Networks, Avraam Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, Alexandros Iosifidis,2017

[7] Support vector machine with adaptive parameters in financial time series forecasting, L. Cao, F. Tay,2003

[8] Cont., R. Empirical properties of asset returns: stylized facts and statistical issues. Quant. Finance 2001, 1, 223–236.

[9] Cheng, J.; Huang, K.; Zheng, Z. towards Better Forecasting by Fusing Near and Distant Future Visions. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI), New York, NY, USA, 7–12 February 2020

