Stock Market Prediction Using Machine Learning

Mr.Tejas Pandav[1], Mr.Chetan Shinde[2], Mr.Malharrao Shelar[3], Mr.Shubham Shelar[4], Mr.Saurabh Kodre[5], Prof. Punam Chavan[6]

[1,2,3,4,5,6]*Department of Information Technology, Zeal College of Engineering and Research, Narhe, Pune, Maharashtra, India*

Abstract -The subject of stock forecasting is currently quite popular. Only a select few individuals had access to the study in the beginning due to a number of factors including the device's limitations. Since science and technology have advanced so quickly, more and more people are now interested in studying prediction, and it is getting simpler and simpler for us to predict stocks using various techniques like machine learning, deep learning, and others. In this essay, we'll use LSTM (Long and Short Term Memory) to predict stocks and incentives for the following day. We can improve pre-processing methods to remove noise from data so that subsequent operations like categorization and prediction doesn't make any impact

Keywords: LSTM, Stock Patterns, Machine Learning, Stock prediction

I INTRODUCTION

I tracked the evolution of return charts over the previous several days, weeks, and months as well as over the course of a single day. In addition, I kept up with current events about the businesses on my shortlist to gain insight into their profitability and potential for performance. When examining the charts itself and taking into account extra information about a particular firm, this strategy really has some similarities to technical analysis. Finally, I made an effort to purchase when the return curve began to decline and sell at higher levels. When I first started investing, this strategy performed admirably. Even though I was a good intraday trader at the time, I ultimately lost a lot of money on the stock market. The fact that I panicked and sold off my whole wealth when the stock markets quickly declined was a major factor. In the end, the prices decreased even further, making this appear like a smart solution. But looking back, I wish I had chosen otherwise because my portfolio would have virtually doubled by this point. At this point, I felt the need to learn more about practical stock market investment strategies in order to perhaps one day make up for the loss.

I was able to focus on the creation of suitable investing strategies and global stock markets during my dissertation assignment. I wanted to understand more about the complex techniques that the academic and commercial worlds use to decide for or against a certain investment plan rather than depending just on intuition. My theory placed a lot of focus on choosing the right moment for selling and buying stocks. I wanted to understand the many strategies used by professionals to boost returns and lower risk in their portfolios and create my own system for wise investment choices.

II LITERATURE REVIEW

In the paper "Stock Graph Pattern recognition with Deep Learning" [1], the effectiveness of CNN and LSTM for identifying typical chart patterns in historical stock data was evaluated. It displays two frequent patterns, the procedure used to create the training set, the designs of the neural networks, and the accuracy levels attained.

In the paper [2], A "window" of different durations is run, and it is divided into frames with durations that scale to one size and values that scale to 1. The received frames are transformed into 2D matrices and fed into a 2D convolutional NN for analysis, which calculates the chance that the frames correspond to the various classes of patterns.

With a convolutional NN reaction time of around 0.65 seconds per 1000 data samples, the detector has an accuracy of about 98.6%, which is consistent with an examination of the closing prices of trades on the exchange for more than 2.5 years.

Traditional predictive regression models suffer considerable difficulties in out-of-sample predictability tests because of model uncertainty and parameter instability, according to a number of studies that offer compelling evidence in support of this claim. Recent research provide specific tactics for resolving these issues. Support Vector Machine (SVM) is a

relatively recent learning technique that uses the kernel approach, has a sparsely distributed solution, and may control the decision function. Using the Support Vector Machines technique to forecast the stock market, we propose a theoretical and empirical framework in this research. First, for additional stock multivariate analysis, four company-specific and six economic variables that might affect the stock movement are chosen. In order to analyse the link between these parameters and forecast stock performance, Support Vector Machine is employed. Our findings imply that SVM is a potent tool for predicting movements in the stock market. [3]

Practically speaking, it is crucial to improve the predictability of models since stock market predictions can result in actual financial gains or losses. As a result, several research have attempted to analyse and predict financial time series using statistical or soft computational techniques that can look at the complex and unstable financial market. Deep learning techniques have been widely used in recent years because to their outstanding results in several categorization issues. In this work, we built an RNN-based stock price prediction model utilising LSTM units, one of the common deep learning techniques. We combined GA and LSTM networks with the model's own architectural components to take the stock market's temporal characteristics into account. The LSTM network utilised in this study has two hidden layers, which is a deep design for more efficiently reflecting the complex and nonlinear characteristics of the stock market. In order to get the ideal or nearly ideal value for the size of the time window and the quantity of LSTM units in an LSTM network, GA was used. [4]

Sentiment analysis has been successful in analysing people's attitudes by looking at enormous social data sets. This is an innovative method to get comments on particular subjects from individuals by depending on social media information. The remaining 30% of the dataset is applied for testing, while the remaining 70% is used for training. The suggested classifier has been shown to accurately classify the positive, negative, and neutral feelings with a 98.32% accuracy rate. It has been determined from the trial that the previous work's sentiment detection accuracy has risen by 8.99%. The addition of a stop word panel to the GUI, allowing a user to add or remove stop words as necessary, is the major benefit of this effort. [5]

When a system uses imagery as inputs and using machine learning techniques to attempt to predict characters from the images provided, the system is said to be offline. In this project, we have focused on the issue of offline character recognition. Over the years, several machine learning techniques have been suggested for tackling this issue. In order to address the offline written character identification problem, we used six of the most well-liked machine learning techniques in this work. We then compared the performance results to determine which technique produced the greatest accuracy results under the given constraints. In the training stage of the chosen machine learning algorithms, we have chosen 92255 photos from the NIST Special 19 Database to be utilised as input images. These techniques include SVM, Decision Tree, Bag of Trees, Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Deep Learning Network with Auto Encoders. We put all of these techniques into practise and compared how well the outcomes performed using the accuracy metric. The comparison's results will be used to assist choose which machine learning technique should be applied to the challenge of offline handwriting character recognition. [6]

Regression architectures may be shown to be able to both capture dynamics and generate predictions. We used stock data to train the model, and it was able to forecast stock price. This shows the capability of the proposed system to find certain connections within the data. Additionally, it is clear from the data that the SVR, RFR, and DTR model is able to detect changes in trends. DTR is determined to be the best model for the suggested technique. It makes predictions using the data available at a specific moment. The DTR model does not outperform the other two models in this case, despite the fact that they are often utilised in other time-dependent data analyses. This is a result of the stock market's frequent unexpected changes. It's possible that changes in the stock market don't always follow a predictable pattern or the same cycle. The existence of trends and their duration will vary depending on businesses and industries. The understanding of these kinds of cycles and patterns will increase investment returns. We have to use networks like DTR to analyse this data since they rely on up-to-date data. [7]

On the basis of the outcomes, we draw the conclusion that there is less connection between the two companies. The stock exchange index has no impact on

the change in stock value. It is based on the opinions expressed on social media. Since there is little correlation between them, predictions made using machine learning algorithms do not produce accurate results. Results are inaccurate because there is less than 50% dependence across all variables. The graph patterns between TCS and Infosys, however, exhibit comparable variance, with the exception of a few places where it was the reverse. To handle such situations, a combination of natural language processing techniques for text analysis and summarization can be used. In addition to the factors taken into study in this study, additional factors, such as inflation, deflation, exchange rates for foreign currencies and gold, international economic policies, etc., may also have an impact on stock prices. Martingales, Mean Reversion, and Momentum are additional methods that may be applied. [8]

This study offers an overview of many approaches, including deep learning, hidden Markov models, ARIMA models, and machine learning. It has been noted that choosing the right parameters for the dataset used for prediction is crucial to achieving high prediction accuracy. A greater rate of accuracy is provided by a variety of machine learning models, as well as by hybrid and ensemble models. Fundamental analysis, which combines feature selection, sentiment analysis, and machine learning and deep learning methods, can be utilised to achieve even higher accuracy. [9]

In the field of finance, predicting stock prices is a crucial topic since it helps with the creation of successful trading techniques. In this study, we offer a general framework for adversarial training for predicting the high-frequency stock market using Long Short-Term Memory (LSTM) and convolutional neural network (CNN). This approach uses the open-access index provided by trading software as input to avoid difficult technical analysis and extensive financial theory study, which offers convenience for the average trader with a non-finance background. In order to examine the impact of the model update cycle on the prediction performance, our study mimics the trading style of an actual trader and employs the rolling partition training set and testing set approach. Extensive testing shows that our suggested method may successfully increase forecast decreases in errors and stock price direction prediction accuracy. [10]

III PROBLEM STATEMENT

The field of machine learning known as pattern recognition focuses on developing various numerical techniques to identify patterns in datasets. Finding patterns in data may also be used to divide it into different categories or predict actions on future datasets. Finding sequences of various sizes and lengths would be made easier by automation. As these signals do have a weak link with prices, it would also help with the provision of useful information for stock market price prediction. According to some research, patterns by themselves are sufficient to forecast trends, but when combined with other variables, they could produce diverse outcomes.

IV PROPOSED SYSTEM

The stock price of a certain commodity and the stock value in previous years serve as training input. From the provided data, statistical characteristics are retrieved, examined, and sent to the classifier for comparison. As training data, a stock chart is created. Commodity stock data are provided as input to the system. A dimensionality reduction technique called feature extraction divides a large amount of raw data into smaller, easier-to-process groupings. These huge databases include a lot of variables, which means processing them takes a lot of computational power.

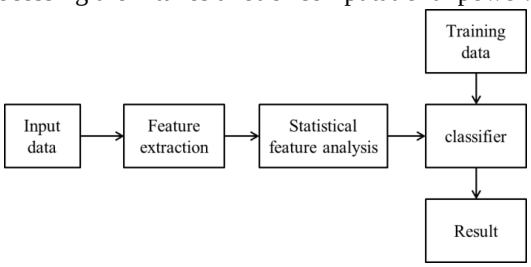

Fig.1: Architecture of proposed system

V ALGORITHM

LSTM

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN). The output from the previous phase is sent into the current step of an RNN as input. Hoch Reiter & Schmidhuber created the LSTM. It addressed the issue of long-term RNN dependency, in which the RNN can predict words from current data but cannot predict words held in long-term memory. RNN's performance becomes less effective as the gap length rises. By default, LSTM may save the

data for a very long time. It is used for time-series data processing, prediction, and classification.

Time series, audio, and text are just a few examples of the sequential data that it is particularly made to handle. LSTM networks are particularly suited for applications like language translation, speech recognition, and time series forecasting because they can learn long-term relationships in sequential data.

When learning long-term dependencies, a standard RNN may find it challenging since there is just one hidden state that is transferred over time. This issue is solved by LSTMs by include memory cells, which are structures that can store data for a long time. Three gates—the input gate, the forget gate, and the output gate—control the memory cell. These gates control what data should be input into the memory cell, removed from it, and output from it.

What data is added to the memory cell is handled by the input gate. What information is deleted from the memory cell is controlled by the forget gate. Additionally, the output gate manages the data that is output from the memory cell. It enables LSTM networks to learn long-term dependencies by selecting which information to keep or discard as it traverses through the network.

LSTM networks, which are capable of identifying even complicated patterns in sequential data, might be made by stacking LSTMs. Convolutional Neural Networks (CNNs) are a type of neural network design that may be used with LSTMs to analyse.

Structure Of LSTM:

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

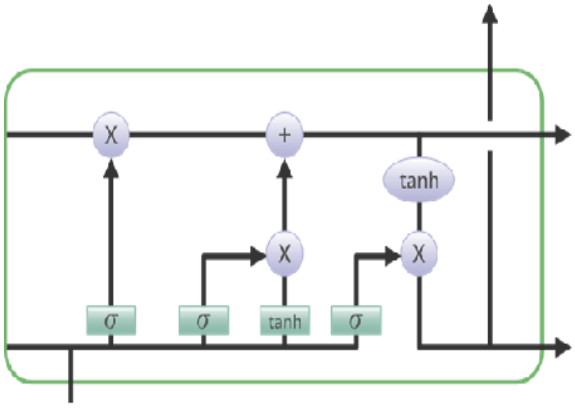

Fig.2 : Architect LSTM

Information is retained by the cells and the memory manipulations are done by the gates. There are three gates –

1.Forget Gate: The forget gate removes information that is no longer relevant in the cell state. The gate receives two inputs, x_t (input at the current time) and h_t-1 (before cell output), which are multiplied with weight matrices before bias is added. The output of the activation function, which takes the outcome, is binary. If a cell state's output is 0, the piece of information is destroyed, however if it is 1, the information is saved for use in the future.

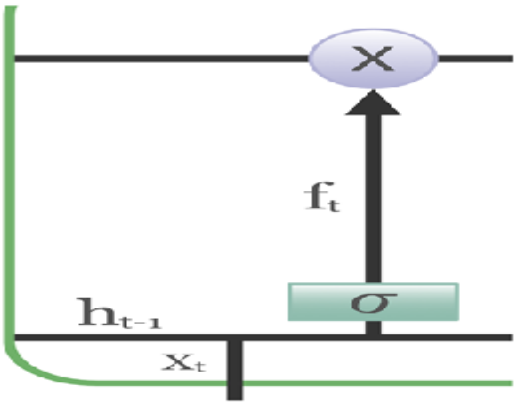

Fig. 3 Forget Gate

2.Output gate: The output gate's task is to collect necessary information from the current cell state and display it as output. The tanh function is first used to the cell to create a vector. The data is then filtered by the values to be remembered using the inputs h_t-1 and x_t, and the information is then controlled using the sigmoid function. The vector's values and the controlled values are finally multiplied and provided as input and output to the following cell, respectively.

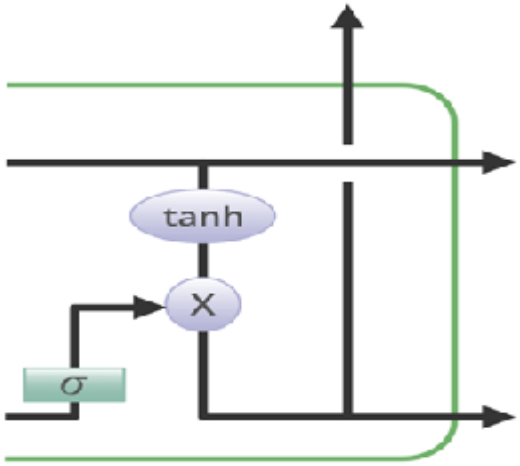

Fig. 4 Output Gate

VI METHODOLOGY

Daily stock market operations include share exchanges and sense calculating. For novice investors looking to invest in the stock market based on numerous criteria such stock data and stock predictions, our prediction technique will be helpful. Based on the stock values of the firm, our programme will carry out crucial stock market tasks and provide investors with improved forecast outcomes. The following different securities exchange criteria are used to segment financial exchange expectations:

The NSE list With the exception of equities that have been suspended for longer than a year, the NSE Index is a composite financial exchange list derived from the prices of every single basic stock on the principal leading body of the Stock Exchange.

Basic analysis The process of doing a fundamental analysis of stocks is helpful in making risky investment decisions. Its primary relevance is in determining a security's intrinsic estimate. It would then be possible to compare it to the current stock price and determine whether or not the stock is overvalued or cheap.

closing the index estimate: The term "closing cost" often refers to the last price at which a stock trades during a regular trading session. Regular trading hours continue for various U.S. marketplaces from 9:30 am to 4:00 pm.

The most basic type of hybrid input is when the cost of an advantage closes on one side of a moving normal and goes from one side to the other. Value hybrids may be used as a crucial exit or passage strategy and are used by merchants to identify changes in the condition of the economy.

looking for stock offers A single offer of many available loads of a company, subordinate, or other financial resource is said to have an offer cost. The stock cost, to put it simply, is the highest price someone is ready to pay for the stock or the lowest price it is possible to find it for.

VII CONCLUSION

Since investors want to know the return on their investments, financial professionals are quite familiar with the technique of anticipating the price of securities. The stock costs are often predicted by professional experts and intermediaries based on historical costs, volume, value designs, and basic trends. The stock value expectation nowadays is more confusing than it used to be since stock prices are affected by an organization's financial situation as well as by the country's economic status, the political climate, and unexpected events, among other things. Since the arrival from the offer market is continually ambiguous and uncertain, traditional methods cannot provide accurate expectations. There has been a great deal of research done in that area, and various budgetary exchanging frameworks for stock value anticipation have also offered advanced perceptive processes that range from pure numerical models and master frameworks to neural systems. In this essay, we'll use CNN (Convolution Neural Network) to predict stocks and a reward for the next day. To remove noise from data and ensure that it has no influence on subsequent operations like classification and prediction, we can improve pre-processing techniques.

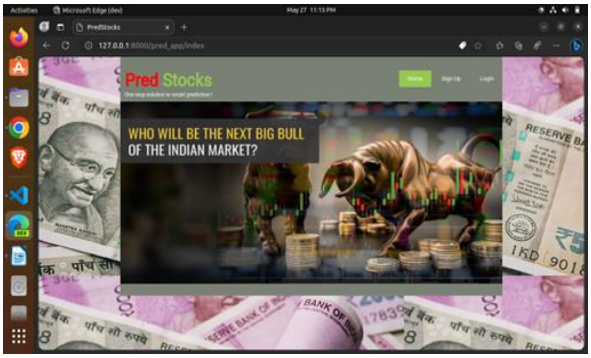

Fig.No 01-Home Page

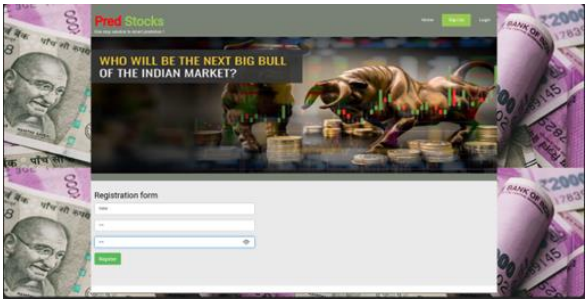

Fig.No 02-Sing Up Page

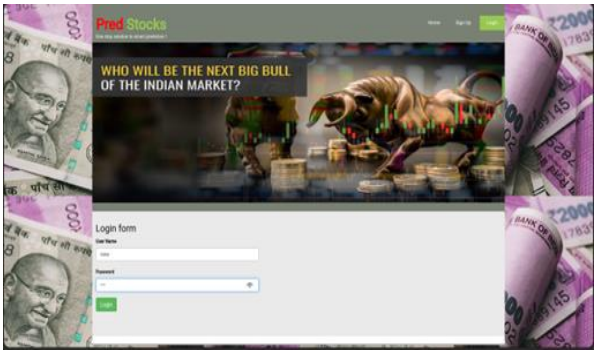

Fig.No 03-Log in Page

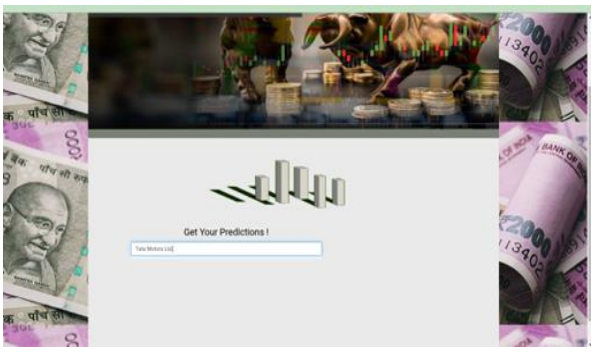

Fig.No 04- Predication Page

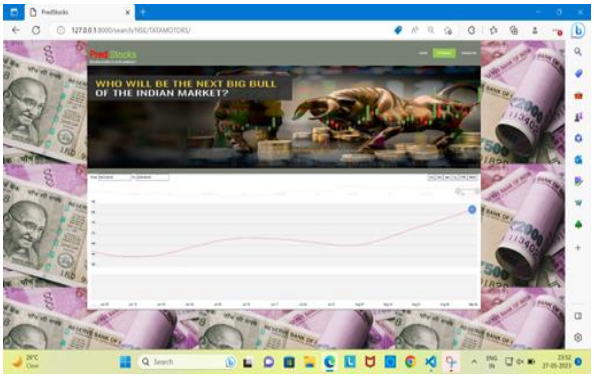

Fig.No 05-Final Result

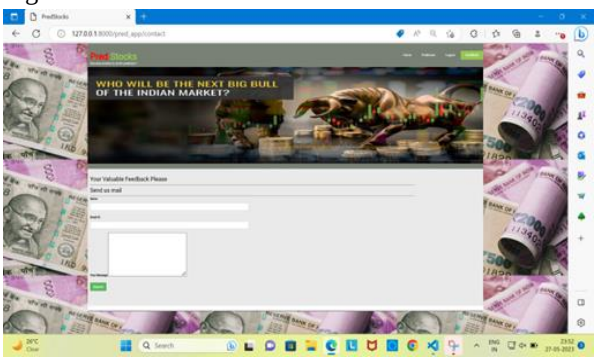

Fig.No 06-Feedback

REFERENCE

[1] Marc Velay and Fabrice Daniel, “Stock Chart Pattern recognition with Deep Learning”, Researchgate, June 2018

[2] Victor Skuratov, Konstantin Kuzmin, Igor Nelin, Mikhail Sedankin, “Application of a convolutional neural network to create a detector of technical analysis figures on exchange quotes charts”, (2019), «EUREKA: Physics and Engineering» Number 6 DOI: 10.21303/2461-4262.2019.001055

[3] International Journal of Current Trends in Engineering & Technology Volume: 02, Issue: 01 (JAN-FAB, 2016) 18 Stock Market Prediction Using Support Vector Machine Zhen Hu 1, Jie Zhu 2, and Ken Tse

[4] Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction Hyejung Chung and Kyung-shik Shin Sustainability 2018, 10, 3765; doi:10.3390/su10103765

[5] An optimized CNN based robust sentiment analysis system on big social data using text polarity feature KomalpreetKaur, ChitenderKaur, TarandeepKaur Bhatia International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-6, April 2019

[6] Performance Comparison of Machine Learning Methods for Solving Handwriting Character Recognition Problem Ş.G.KIVANÇ1, A.E. BAKTIR2 and B.ŞEN3 International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES’18), May 11-13, 2018 Safranbolu, Turkey

[7] JSRD - International Journal for Scientific Research & Development| Vol. 6, Issue 07, 2018 | ISSN (online): 2321-0613 All rights reserved by www.ijsrd.com 408 Stock Market Prediction using RFR, DTR & SVR Ravikant1 Suman Kumar Swarnkar2L. P. Bhaiya

[8] International Journal of Computer Sciences and Engineering Open Access Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693 Stock Market Analysis and Prediction using Hadoop and Machine Learning Piyush Jain, KaustubhBhat, HarshalKesharwani3, PriteshBhate, Khushboo P Khurana

[9] International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072 © 2019, IRJET | Impact Factor value: 7.211 | ISO 9001:2008 Certified Journal | Page 4842 Stock Market Forecasting Techniques: A Survey Rashmi Sutkatti1, Dr. D. A. Torse

[10] Xingyu Zhou, Zhisong Pan, Guyu Hu, Siqi Tang, Cheng Zhao “Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets”, Mathematical Problems in Engineering Volume 2018, Article ID 4907423, 11 pages.