

# Stock Market Prediction using Machine Learning Algorithms

**Dr. K. Velmurugan<sup>1</sup>, Meera. T<sup>2</sup>, Meenatchi. R<sup>3</sup>**

Head, Department of Computer Science and Engineering<sup>1</sup>

Students, Department of Computer Science and Engineering<sup>2,3</sup>

Anjalai Ammal Mahalingam Engineering College, Kovilvenni, Tiruvarur, Tamil Nadu, India

**Abstract:** Machine learning is effectively implemented in forecasting the stock prices. The objective is to predict the stock prices in order to make more informed and accurate investment decisions. The essence of the issue lies in exactly forecasting the future stock price of the reputed firm, based on historical or past prices, using Recurrent Neural Network algorithm, LSTM-LongShort Term Memory, ANN-Artificial Neural Network, CNN - Convolutional Neural Network. We also used regression as a trial and error method and found the limitations in the case of a non-continuous dataset. The findings of the study would lend a hand to a common investor averting huge losses and optimizing the stockholder investing in beneficial stocks. While predicting the actual prices of a stock is an uphill climb, we can build a model that will predict whether the price will go up or down.

**Keywords:** AI, ML, Neural networks, LSTM, Moving Average, Linear regression, CNN, ANN.

## I. INTRODUCTION

A correct prediction of stocks can lead to huge profits for the seller and the broker. Frequently, it is brought out that prediction is chaotic rather than random, which means it can be predicted by carefully analyzing the history of respective stock market. Machine learning is an efficient way to represent such processes. It predicts a market value close to the tangible value, thereby increasing the accuracy scenarios and variables, allowing them to make quick investment decisions. All of this implies that there will be a large amount of data to analyze. As a result, financial analysts, researchers, and data scientists seek to figure out new ways to discover the stock market trends using analytics. This on Intelligent Trading Systems to predict prices based on gave rise to the concept of algorithmic trading, which involves the implementation of orders using automated, pre-programmed trading techniques. As one of the most important statistical data, the closing value of the stock index has been applied to derive helpful information about the current and likely future movement pattern of stock markets. This study used a predictive analytical approach to accurately estimate the future stock price of business stocks and discover the best fit model for algorithmic trading.

## II. LITERATURE REVIEW

Significant work has been done throughout the world in this field. A testament to which is the work of M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali [1] and that of K. Raza [2] who have surveyed the application of machine learning techniques and presented the current advancements in this field.

Vaishnavi Gururaj et al. [2] Predicting the stock price using the learning Approaches for algorithms of Linear Regression and Support Vector Machines. The methods in this paper include the segment of Environment(R), Time-Series Forecasting, and Sliding-window method.

Sioiokwong et al.[3] Stock price forecasting is based on short-term patterns. Using KNN, it is demonstrated that the upgraded KNN model performs better and more precisely than logistic regression analysis and the classic KNN model in terms of prediction.

Hyeong Kyu Choi, B.A[4] Student Dept. of Business Administration Korea University Seoul, Korea Stock Price Correlation Coefficient Prediction with ARIMA-LSTM Hybrid Model

Parshvchhajer et al. [5] Artificial Neural Network, Support Vector Machine, and Long-Short Term Memory are used to forecast stock prices. Before the introduction of Algorithms, statistical methods such as ARIMA, ESN (Echo State Networks), and regression were utilized in this paper. These strategies were applied in both a linear and a singular manner.

### **2.1 Need of the Study**

Businesses primarily run over customer's satisfaction, customer reviews about their products. Shifts in sentiment on social media have been shown to correlate with shifts in stock markets. Identifying customer grievances there by resolving them leads to customer satisfaction as well as trustworthiness of an organization. A trade-off between risk and return, as well as sensitivity to stock price swings, is necessary to maximize the returns on stock investments. This research will assist a broad range of investors (retail investors, financial firms, stock funds, private lenders, and overseas institutional investors) in making well-informed investment selections based on scientific thinking and logical decision-making. You know you need insights to inform your decision making and you know that you're lacking them, but don't know how best to get them. Sentiment analysis provides some answers into what the most important issues are, from the perspective of customers, at least. Because sentiment analysis can be automated, decisions can be made based on a significant amount of data rather than plain intuition.

### **2.2 Problem Statement**

Time Series forecasting & modelling plays an important role in data analysis. Time series analysis is a specialized branch of statistics used extensively in fields such as Econometrics & Operation Research. Time Series is being widely used in analytics & data science. Stock prices are volatile in nature and price depends on various factors. The main aim of this project is to predict stock prices using Long short term memory(LSTM), MovingAverage, ARIMA, KNN, LinearRegression. Market knowledge and personal finance are the key aspects for investors to consider while making investment selections. Lack of these abilities among stockholders would result in inconsistency and inaccuracy in market predictions, resulting in investors investment losses. The financial sector builds and offers stock price forecasting strategies for investors. However, investors have no established forecast technique, which exacerbates the scope and severity of the problem. As a consequence, this research was done.

### **2.3 Machine Learning for Financial Instruments**

ML has been used in a variety of sectors, including finance and economics, in recent years. Many researchers have employed machine learning algorithms to develop tools for analyzing historical financial data and other associated data (such as economic conditions) to aid investment decision-making. By combining business news and social media data with machine learning algorithms, a promising non-linear prediction model was developed to estimate the stock values of construction companies. More crucially, correct results need the use of historical or time-series financial data, as well as the careful selection of relevant models, data, and features. The accuracy of the outcomes is determined by the infrastructure used, the data collected, and the algorithms used. The more accurate the ML result, the higher the data quality.

### **2.4 Computer Learning for Stock Price**

The role of stock price forecasting has a crucial part of investments since accurate stock price predictions can offer trading strategy recommendations. Due to the uncertainty in the future, Stock price predictions based on previous data are not guaranteed to be accurate. Stock prices, for example, are subject to political and economic fluctuations. As a response, for stock price prediction, investors have combined fundamental and technical analysis. The fundamental analysis goes on the way to determining a stock's inherent worth by assessing numerous internal as well as external factors that may affect the stock or company's value. The economic environment, financial results, economic data, and political and social behavior are some of the most important elements. Financial analysis is a technique that uses historical data to forecast future stock prices. This strategy examines trends in the commodities prices such as the daily beginning, peak, lower, and shutting prices. Other characteristics could be researched and used in the technical indicators to improve prediction sureness.

- The initial price of any publicly traded stock at the start of a trading day on an exchange is known as the opening price.
- Peak and Lower Prices - stock's maximum and minimum prices upon this particular day. Traders typically use these statistics to determine the stock's volatility.

- Close Price- stock's price at the conclusion of the day's business.
- Volume- total securities or obligations amount traded in all marketplaces for trading over a specific period.
- Modified Ended Prices- genuine cost of a stock; it indicates the potential investment after rewards have been distributed.

Many algorithms and strategies for stock price prediction have been suggested in the literature, and some of them summarise the functionality of different classifiers with approaches reported with research. Many deep learn approaches such as RNN, LSTM, Linear Regression performed well in terms of creating low error percentages.

### **III. METHODOLOGY**

Our goal is to forecast future prices and calculate the company's potential growth over various periods. The prediction error for each company in each industry is then examined. We can then determine which time is ideal for predicting the future of that particular industry. With the use of LSTM, anticipate the long-term target price of a company against a set of pre-determined zone. The future prediction will be based on historical data, and we will calculate the progress of those companies over various time. Then, after assessing the targeting price variances on every time interval, we choose the schedule with the highest growth, i.e. the least inaccuracy for the given region. Let's pretend we're looking at data for Years in our analysis.

#### **IMPLEMENTATION STEPS:**

- Step 1: Raw Stock Price Dataset: The finance yahoo official website is used to collect day-by-day past stock prices of selected firms. Load the training dataset
- Step 2: Import libraries
- Step 3: Pre-processing: Data analysis strategy for refining data into usable and productive information.
- To evaluate, the data collected is divided toward training and testing sets once it has been turned into a clean dataset. Making a data structure with 30-time steps and only one output.
- Step 4: Feature Selection: In this step, the data features that will be supplied to the Neural network are chosen. Date and End Price were assigned as variable extracted in this investigation.
- Step 5: Fitting the model using Linear Regression, CNN, ANN.
- Step 6: Plotting the actual and predicted prices and understanding the limitations of regression.
- Step 7: Conversion of the dataset into a time series object.
- Step 8: Set the Neural Network Imitation: The training dataset is accustomed to train NN pattern. Random weights and biases are used to start the model. The proposed LSTM model consists of a serial input, three LSTM levels with hidden layers with stimulation. The output layer is once again a serous membrane with a linear perceptron.
- Step 9: Generate Output: The error difference between the RNN-generated output and the target values is determined. By modifying the biases and weights, the Back propagation method is utilized to reduce the error difference.
- Step 10: comparing accuracy for ANN, CNN, and LSTM and conclude the best fit model for our Concluding with the best accurate model

Prior to the implementation of the algorithm:

A positive correlation indicates that there exists a direct relationship, and the data is scaled using a mini max scalar to a fixed range usually 0 to -1. splitting the data 80% train, 20% test.

Correlation:

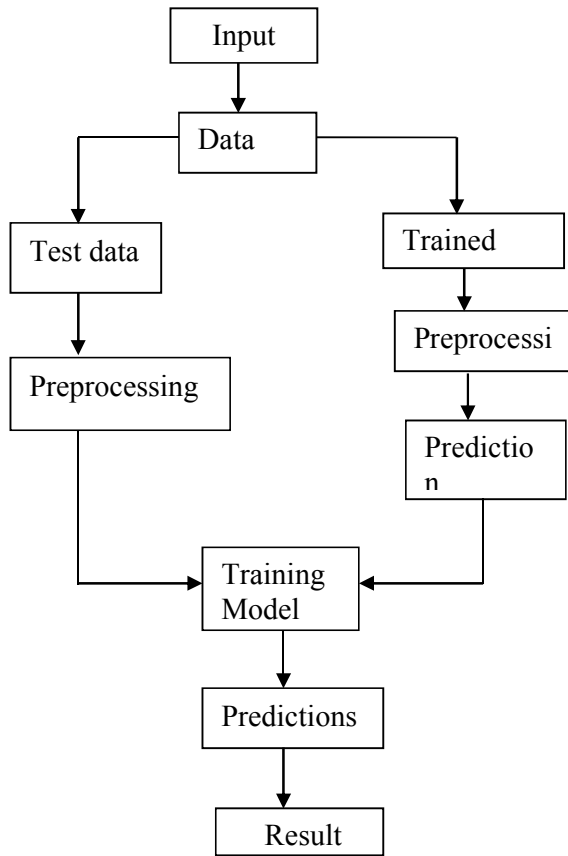
(i) Correlation assists in determining the degree to which variables are related. It gives us a chance to make decisions about our future events.

(ii) Correlation analysis assists in making the nature and scope of a relation, which can be utilized to plan and forecast future events.

After implementation of the algorithms,

**Accuracy test:**

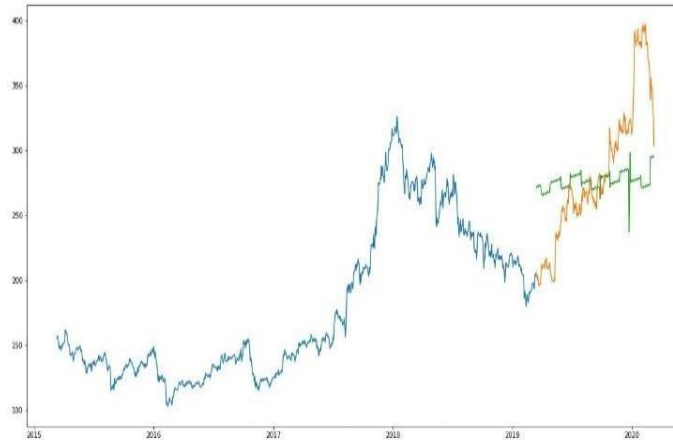
Accuracy is generally the no. of correct predictions to the total predictions. By visualizing through sea born we could able to identify that all the other variables have a linear relationship with close. So, linear regression is performed for prediction. The efficiency of a system is linked to the profits obtained by implementing it.



**IV. TRAINING AND PREDICTION**

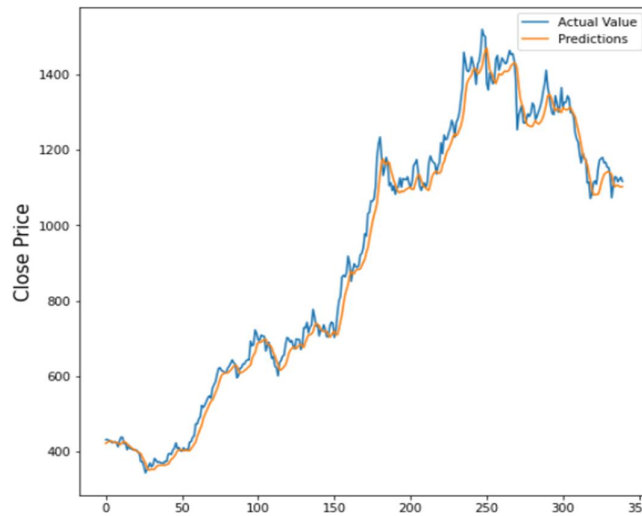
**4.1 Linear Regression**

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable. The regression equation is solved to provide the coefficients, which are then used to forecast a stock's future price. The biggest boon of linear regression models is linearity: It makes the estimation procedure simple. Linear regression models can also be considered single-layer neural networks because they are made up of just one artificial neuron. We can consider this transformation as linear regression because every input is associated to every output. But here for this data, linear regression cannot be fitted because the volume variable is highly correlated and that collapse the total system. so, we are converting the raw data to a time series object.



#### 4.2 ANN

Artificial Neural Networks (ANNs) are a class of computational algorithms that takes influence from the cerebral cortex. That's also, much as a neuron in the structure of the brain can learn from prior data, artificial neural networks may discover insights and replicate these in the shape of suggestions or categorization. The accuracy of this model for our dataset is 96.05%



#### 4.3 CNN

A feed-forward neural network is a convolutional neural network. Convolutional neural networks include the same properties as standard neural networks, including intake layers, screened layers, and outcome layers, with the return of the former layer of inversion or pools being the supply of the level of convolution. Pooling layers, fully-connected strands, and other features distinguish them. In a convolutional neural network, the number of levels is higher than in a regular cnn architecture, demonstrating the neural network's capability to some extent. The bigger the number of hidden layers, the more information it can extract and recognize from the input. The accuracy for this model is 98.79%



#### 4.4 RNN

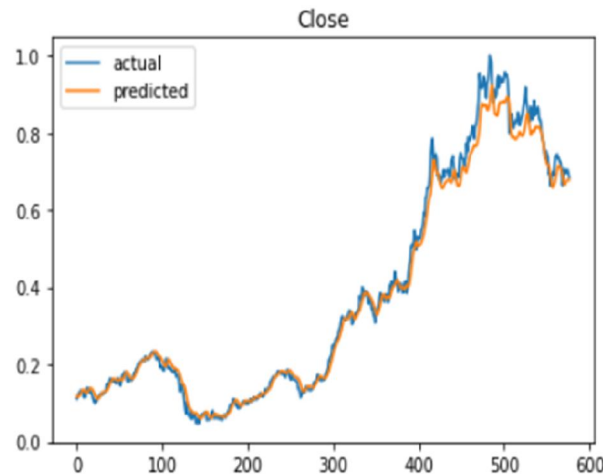
RNNs are a learning algorithm that is efficient and are the most intriguing procedures with local storage. As a result, a Recurrent neural network incorporates two inputs: the current and recent past. This is critical because the data sequence offers critical information about what will happen next, which is why an RNN can perform tasks that other algorithms lack. It analyses prolonged sequences. This makes RNN ideal for stock price prediction, which requires a large amount of previous data. LSTM can indeed be thought of as an RNN with a computer's memory and two major trajectories:

1. In a narrow condition, the rates remain at the present moment step.
2. Protracted state, while moving across the network, saves, examines, and refuses objects intended again for brief.

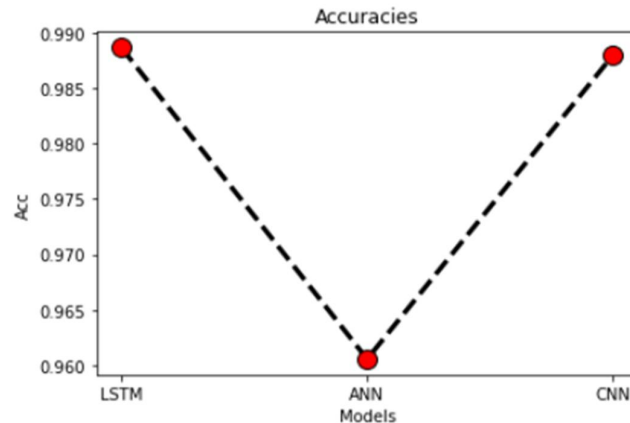
#### Long Short Term Memory (LSTM):

A sort of recurrent neural network is the LSTM. It was created specifically to address the general RNN's long-term dependency problem. It's been applied in a variety of fields, including language processing, voice search, visual caption creation, film labeling, economic analysis and language processing. A chain form of the repeated neural network module is seen in every RNN.

Thus, it is highly suited to learning from significant experiences. We could obtain an accuracy of 98.86%.



Comparison of accuracy of LSTM, ANN, and CNN after converting it data time series object:



Why LSTM is concluded as the best model for stock price prediction?

Models are most commonly used with univariate data.

Long spans of time divide them. The units of an LSTM are used to build the stages of an RNN, also known as an Lstm model. Because of Long short - term memory, RNN can remember inputs for a long time. Because LSTMs store data in an address space comparable to that of a processor, this is the case. The LSTM can acquire, copy, and remove content out of its memories.

Although significant incidents in a series data may have unpredictable lags, LSTM networks are more for diagnosing, analyzing, and drawing conclusions.

## V. CONCLUSION

In this project, we are predicting closing stock price of any given organization, we developed a web application for predicting close stock price using LMS and LSTM algorithms for prediction. We have created and evaluated that model using TATA Steel Product past data from January 1, 2015, to December 30, 2021. The source data is appropriately pre-processed, and appropriate factors for creating predictive models are found. The prediction framework is further enhanced by developing three Convolution Neural Network methods plus varied input data sizes and network configurations following the creation and validation of cognitive computing models. These CNN-based layouts outperformed machine-learning-based predictive models by a wide margin. The study has convincingly proven that deep learning-based models are far more able to extract and learn the characteristics of a training dataset than their machine learning alternatives. we could achieve the accuracy of 98.86% and concluding LSTM is the best performing model for the dataset we took .though we could get high accuracy time series object has its limitations, unlike other algorithms it won't support missing values, data transformations are costly and it demands linear relationship between variables.

## REFERENCES

- [1]. Adil MOGHAR, Mhamed HAMICHE, "Stock Market Prediction Using LSTM Recurrent Neural Network "International Workshop on Statistical Methods and Artificial intelligence (IWSMAI), April 6-9,2020,warsaw,Poland.
- [2]. Emerson, S., Kennedy, R., O'Shea, L., & O'Brien, J. (2019, May). Trends and Applications of Machine Learning in Quantitative Finance. In 8th International Conference on Economics and Finance Research (ICEFR 2019).
- [3]. VaishnaviGururaj , Shriya V R, Dr.Ashwini K "Stock Market Prediction using Linear Regression and Support Vector Machines" in International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 8 (2019) pp. 1931-1934
- [4]. P. Abinaya, V. S. Kumar, P. Balasubramanian, and V. K. Menon, "Measuring stock price and trading volume causality among nifty50 stocks: The todayamamoto method," in Advances in Computing, Communications, and Informatics (ICACCI), 2016 International Conference on. IEEE, 2016, pp. 1886–1890.
- [5]. Moritz, B., & Zimmermann, T. (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. Available at SSRN 2740751.

- [6]. Pramod, Mallikarjunashastry Pm, " Stock Price Prediction Using LSTM ".Test Engineering and Management 83(May-June 2020):5246-5251, January 2021
- [7]. RezaulKarim, MdKhorshedAlam, MdRezaulHossain, "Stock Market Analysis Using Linear Regression and Decision Tree Regression" in 1st International Conference on Emerging Smart Technologies and Application, August 2021.
- [8]. "Stock price prediction using LSTM, RNN, and CNN-sliding window model" in International Conference on Advances in Computing, Communications and Informatics (ICACCI), September 2017
- [9]. S. Kavitha, S. Varuna, and R. Ramya, "A Comparative Analysis on Linear Regression and Support Vector Regression," in Online International Conference on Green Engineering and Technologies (IC-GET), 2016.
- [10]. Sioiok Wong, "Stock Price Prediction Model Based on the Short-term Trending of KNN Method" in 7th International Conference on Information Science and Control Engineering (ICISCE), 20 September 2021.
- [11]. KimberlyAmadeo, How the stock market works [internet], 2021, Available from: <https://www.thebalance.com/how-does-the-stock-market-work-3306244> [cited 2021 Jun 25].
- [12]. Dharmaraja Selvamuthu, Vineet Kumar and Abhishek Mishra Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India - Indian stock market prediction using artificial neural networks on tick data.
- [13]. Lufuno Ronald Marwala A dissertation submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Engineering - Forecasting the Stock Market Index Using Artificial Intelligence Techniques.
- [14]. Xiao Ding, Kuo Liao, Ting Liu, Zhongyang Li, Junwen Duan Research Centre for Social Computing and Information Retrieval Harbin Institute of Technology, China - Event Representation Learning Enhanced with External Common-sense Knowledge.
- [15]. Huicheng Liu Department of Electrical and Computer Engineering Queen's University, Canada - Leveraging Financial News for Stock Trend Prediction with Attention-Based Recurrent Neural Network.
- [16]. Hyeong Kyu Choi, B.A Student Dept. of Business Administration Korea University Seoul, Korea = Stock Price Correlation Coefficient Prediction with ARIMA-LSTM Hybrid Model.
- [17]. M. Nabipour Faculty of Mechanical Engineering, Tarbiat Modares University, 14115-143 Tehran, Iran;Mojtaba.nabipour@modares.ac.ir - Deep Learning for Stock Market prediction.
- [18]. Lavanya Ra SRM Institute of Science and Technology | SRM-Department of Computer Science-Stock Market Prediction.
- [19]. M. Nabipour Faculty of Mechanical Engineering, Tarbiat Modares University, 14115-143 Tehran, Iran; Mojtaba.nabipour@modares.ac.ir - Deep Learning for Stock Market prediction.
- [20]. Lavanya Ra SRM Institute of Science and Technology | SRM · Department of Computer Science - Stock Market Prediction.
- [21]. M. Mekayel Anik · M. Shamsul Arefin (B) Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh - An Intelligent Technique for Stock Market Prediction.
- [22]. K.A.Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, 2018, pp. 151-156