ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 11, Issue 4, April 2024

Stock Price Prediction Model Using Machine Learning

[1] Deepanshu Singh, [2] Sahil Kuamr, [3] Vandana Choudhary, [4] Namita Goyal

[1] [2] Student, Department of IT, Maharaja Agrasen Institute of Technology, Delhi, India
[3] Assistant Professor, Department of IT, Maharaja Agrasen Institute of Technology, Delhi, India
[4] Assistant Professor, Department of IT, Maharaja Agrasen Institute of Technology, Delhi, India

***Abstract*— This article provides an exhaustive analysis of stock price prediction models using deep learning based on LSTM neural networks. An advanced sub-type of RNN known as LSTM has risen in importance within financial forecasting due to its ability to detect long-term dependencies and patterns of time series data. The paper discusses different aspects of LSTM-Stock price prediction models including architecture designs, data pre-processing, feature engineering, and evaluation measures. The study evaluates various algorithms including Logistic Regression, SVM, Linear Regression, & LSTM. Additionally, it reveals the issues involving LSTM models when attempting to forecast share price movements. This research adds to the growing literature on utilization of deep learning in predicting stock prices, and can be useful to both scholars and practitioners interested in financial prediction. These outcomes suggest the strength of deep learning methods as powerful means of enhancing decision making in such volatile and uncertained area as stock market.**

***Index Terms*— stock price forecasting, machine learning, LSTM, linear regression, Logistic regression, SVM.**

I. INTRODUCTION

The stock market is a marketplace where people trade stocks, also known as shares, from public companies. Predictions of stock prices are important particularly in finance and it leads to increased interest in applying machine learning methods to predict them accurately. Due to their capacity to construct serial information, LSTM networks are effective tools for this line of work. In an effort to design a stock data predictor program powered by LSTM models, this study aims at forecasting the prices of the shares of a certain company. Conventionally, stock trend forecasting algorithms make use of time series of stock prices which are statistically analyzed. A new technique of employing the deep learning based LSTM network models on historic values of given share to project future cost trend. This paper will focus on how to use the stock price dataset for Infosys whose headquarters are located in Bangalore (India), one of the leading companies in global consulting and the IT sector.

The price of stock for Infosys was volatile from 2015 through the year 2020. These swings can be attributed to a number of factors, including industry competition, legal amendments, or the company's fiscal performance. This sector underwent some developments that involved emergence of new competitors and change of regulatory requirements. As result, investors followed closely Infosys's financial reports, number of subscribers and plans to move with the changes taking place in the consulting industry. Infosys' stock price might have been influenced by the condition of the world economy or the ongoing global politics. Various studies based on recurrent neural network, for example, LTSM network, have provided promising results towards the solution of linear problem[1]. They use long term memory unit that enables them to learn from long duration sequences, making them more resilient compared to other networking structures.

II. LITERATURE REVIEW

Stock price forecast is one of the important areas in financial market research and has been attracting much attention from both researchers and practitioners. These complex and volatile market conditions underlie the efforts towards developing exact and robust models that attempt to predict stock prices at any given time. Numerous researches have considered a variety if different methods including classical statistics and modern machine learning techniques in an attempt to uncover signals which can make a prediction for price fluctuations. Technical indicators, fundamental analysis, sentiment analysis, and the addition of macroeconomic data are a few examples that scholars have been looking into to improve the accuracy of their predictions. However, continuing problems remain such as the non-linearity of market behavior, the effects of unpredictable events and volatile mood shifts. Recently, many studies related to advanced prediction models that can understand complex patterns and interconnections of financial time series data has been raised because of progresses in artificial intelligence and deep learning[2]. Although great strides have been made with respect to stock price predictions models, the quest for a uniformly reliable one that is relevant with respect to the dynamics presented in financial markets persists as researchers try to combine different techniques while being inter disciplinary in their approaches.

III. RELATED WORK

The overview considers information pertaining to the current stock market forecasting frameworks because of this. Stock return decision has been a vital research area in the last 20 years. However, empirical evidence has shown that the slope of financial trade balance returns is non-linear [3], which proves the need for a break with linear assumptions. There are several means with which nonlinear and quantified estimations for stock prices could be achieved. However, it was necessary to develop nonlinear models prior to implementing any valuation. Without a doubt, one would accept that financial transactions occur in an environment full of noise, uncertainty, chaos and disorder. Different methods can be used to define boundaries. Such as integrals, double boundaries, straight boundaries, elongated sigmoids, brown. You talked about your studies conducted in the use of AI methods for predicting financial markets. The subject of stock trading speculation is becoming increasingly relevant.Q One of them is concrete evaluations, which have however unclear and non-reliable results. This means that emphasizing more on dynamic rightness testing is important. In the respect of different accessories, every apostate system has its own specific favorable conditions and hindrances. However, for example, this Strait renegade could be applied to eliminate bad fits and failures associated with other standards. It can be adjusted. Error [4]. You can again use least squares to fit a nonlinear model. A developing model involves professional research on the influence of monetary ratios and estimation of inventory costs by employing irregular hinterlands, application of artificial intelligence and artificial mindfulness concepts for anticipating inventory cost.

IV. METHODOLOGY

Data Collection

At this point, we can move on to predicting the Infosys stock price by taking into account historic stock figures of reliable origin. There exist trusted sources like financial databases, API's including Yahoo finance and NSE INDIA and the financial news sites.[5] The relevant characteristics entail opening, closing price, high, low price, trading volume, and other metrics that may influence or determine the movement of price.

Data Preprocessing

After the data collection, attention to missing data and outliers must follow suit. This makes the data valid for subsequent use. It is also necessary to transform time series data to the appropriate format for machine learning methods. Another aspect that needs to be done is normalizing or scaling of the data so that all features appear within similar range [5]. This makes sure that some of the scales do not dictate the analysis because they are large.

Feature Engineering

Addition of extra characteristics enhances predictability of model. Such characteristics are meant to improve your ability to detect the trends and patterns in the stock prices properly. Example involves using sentiments for moving average, RSI, MACD, etc. regarding Infosys with financial news as well. Inclusion of lag facilities which are the past values of target variables can assist in discovering patterns of stock price movements [6].

Splitting the Data

After preprocessing and feature engineering, it is necessary to partition the dataset into training and testing sets. The model is trained with the training set, and the test set is used to evaluate performance. Time-based classification is often used to simulate real-world situations in which a model is trained on historical data and tested on more recent data. This classification allows the model to learn from the past and make predictions about future data.

Model Selection

It is crucial to select the right device getting to know set of rules for time collection forecasting. Several options are available: Examples encompass deep learning strategies inclusive of linear regression, selection trees, random forests, help vector machines, and even Long Short-Term Memory (LSTM) networks. Each set of rules has its own strengths and considerations, and it's miles crucial to pick out the only that best suits the particular problem of predicting Infosys stock rate.

Model Training

Once you select the model, you can train the model with the training data set. Tuning of hyperparameters is important to improve model performance. Superparameters are parameters that describe model behavior, and their optimization can have a significant impact on prediction accuracy. Time series cross-validation methods can be used to assess model performance and ensure generalizability to unknown data.

Model Evaluation

Once a model is trained, it is important to test its performance on a set of test data. Common metrics used for analysis include mean absolute error (MAE), root mean square error (MSE), and root mean square error (RMSE). These metrics provide information about how well the model's predictions match the actual stock price. Evaluating the performance of your model will help you identify areas for improvement.

Hyperparameter Tuning

To similarly optimize model performance, you want to excellent-tune the hyperparameters. This procedure can also encompass techniques together with grid search or random

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 4, April 2024

seek to discover the pleasant mixture of hyperparameters. Fine-tuning the model will allow it to better seize the underlying patterns and traits in Infosys' inventory charge.

Prediction

Once the model is educated, evaluated, and delicate, it is able to be used to expect future inventory fees. By inputting new information into the version, predictions may be made and provide insight into the destiny performance of the Infosys inventory rate. However, it is vital to word that stock market predictions are subject to diverse uncertainties and external factors that could affect their accuracy.

Monitoring and Updating

Building a version and issuing a prediction is most effective step one inside the inventory price prediction process. It is vital to maintain an eye on the version's performance for you to replace it as new records becomes to be had. This would possibly increase the version's forecast accuracy and permit it to alter to moving market situations. Retraining the version on a ordinary basis ensures that it remains applicable and can adjust to the ever-changing inventory marketplace.

Risk Management

Predicting stock prices calls for the utility of threat management strategies. Because of the inventory market's inherent volatility, investing in it could be volatile. As a result, it is essential to understand the dangers worried in inventory fee forecasting and to create plans to reduce them. These strategies can encompass diversifying funding portfolios, putting prevent-loss orders, and applying chance control techniques to protect in opposition to surprising market movements.

V. LINEAR REGRESSION

Forecasting stock prices requires risk management techniques. Because of the inherent volatility of the stock market, investing in it can be risky. Consequently, it is important to understand the risks associated with stock price forecasting and develop policies to mitigate them. To hedge against unexpected market fluctuations, these strategies may include the implementation of risk management policies, stop-loss mandates, and diversification of financial structures.

VI. SUPPORT VECTOR MACHINE (SVM)

The use of machine support vector (SVM) for stock return expectations involves the use of directed learning statistics to spin verifiable stock information and make expectations about future stock returns SVM provides special value clustering and iterative functions, making it suitable for predicting whether a stock price will rise or fall. The statistical foundation of SVM is built on the standard factual learning concept, which emphasizes search selection constraints that limit the common errors in hidden information SVMs use a hyperplane, selecting higher boundaries, giving rise to edges between two classes is large. The edge is the distance between each square of the hyperplane and the neighboring most critical points. Increasing the edges improves the computational ability of the SVM, i. it can work well on other cases for which it is not prepared.

VII. LOGISTIC REGRESSION

Logistic regression is a supervised system learning algorithm used for binary class tasks. It predicts the opportunity that an instance belongs to one of the instructions, making it a probabilistic classification set of rules. Unlike linear regression, which predicts continuous values, logistic regression predicts discrete values between zero and 1. This makes it appropriate for duties which includes unsolicited mail filtering, email type, and fraud detection. The logistic regression equation is given through x) = 1 / (1 e^(-z)) wherein:

- y is the binary elegance label (0 or 1)
- x is the input characteristic vector.
- z is a linear combination of enter features and their corresponding weights.
- e is the bottom of the natural logarithm.

The logistic regression equation can be interpreted because the opportunity that an instance belongs to the high-quality magnificence (y = 1) given the input functions (x). The sigmoid function (1 / (1 e^(-z))) compresses the output of a linear aggregate into a price between 0 and 1 that represents a opportunity.

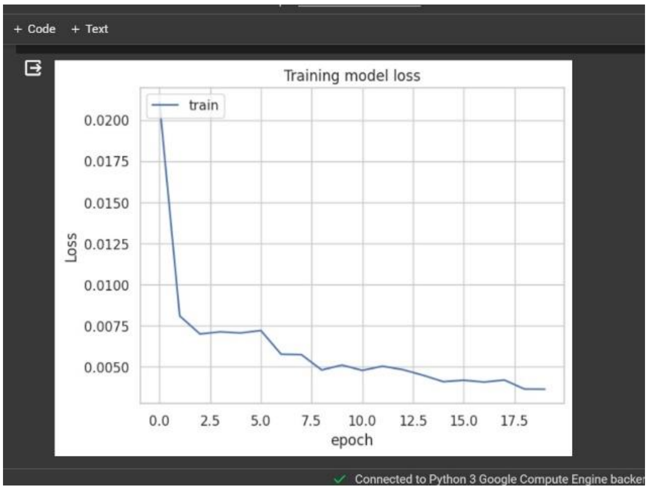
Training the Logistic Regression Model

The purpose of training a logistic regression model is to find the weights that minimize the loss function. The most common loss function for logistic regression is the binary cross-entropy loss, which considers the difference between the predicted probabilities and the actual squared labels [7 Loss = -∑[y * log(P(y = 1 |). x)) + (1 - y) * log(1 - P(y = 1 | x))] Where:

- y is the actual class name (0 or 1).
- P(y = 1 | x) is the predicted probability using a logistic regression equation.

Evaluation of logistic regression models

Once a logistic regression model is trained, you can evaluate its performance using various metrics such as accuracy, precision, recall, and F1 scores. These metrics measure the ability of the model to correctly classify instances into positive and negative.



VIII. RANDOM FOREST CLASSIFIER

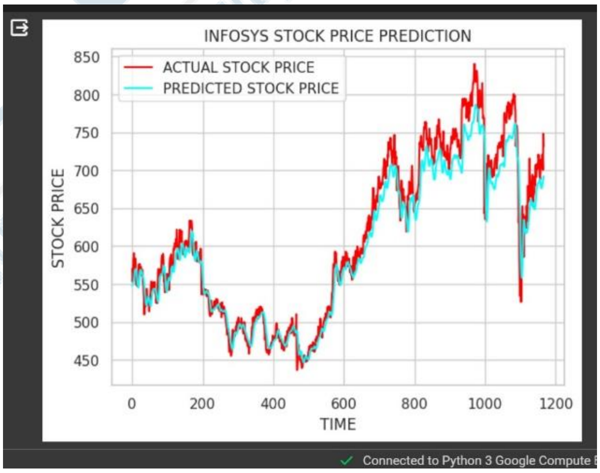
Random forest is a supervised machine learning algorithm that predicts by connecting multiple decision trees. It is a cluster learning method that uses the collective intelligence of different models to improve the overall accuracy of the forecast. Random forests are widely used for classification and regression applications and are known for their efficiency in handling the strong relationships between features and target features Random forests work by on a number of decision trees generated during training. Each decision tree is trained on a different subset of data and random subsets. This process, known as bagging (bootstrap aggregation), helps to reduce correlations between trees and prevents too many unsuitable trees. Builds predictions of all decision trees in the random forest forest to determine a new data point. In the classification task, the class receiving the most votes in the trees is chosen as the predictor. In the regression function, the average predicted values of all trees are used as predictors.

Date	Open	High	Low	Close	Adj Close	Volume
01-07-2015	494.5	502.5	493	498.7	415.5612	6880852
02-07-2015	499.5	500.7	492.525	494	411.6447	4007568
03-07-2015	494	496.5	491	495.15	412.6031	2695306
06-07-2015	492.5	494	487.5	491.65	409.6865	4305602
07-07-2015	492.5	495	489.5	490.25	408.5199	3497418
08-07-2015	489.5	489.5	477.4	478.75	398.9371	7024178
09-07-2015	478.75	483	467.025	469	390.8125	8587772
10-07-2015	474.05	476.05	466.325	468.75	390.6042	7411522
13-07-2015	470.5	477.5	467.025	475.075	395.8748	6531768

IX. PROPOSED SYSTEM

As discussed in the previous section, verifiable information from the market is an important step. Next, we follow the proposed framework that requires separating the intended data analysis items, dividing them into readiness test data, preparing estimates for cost prediction, and providing data set as the final step of the process.

The standard LSTM unit consists of a mobile phone, a data input, an input and a visual input. The cell stores values at intervals, and the three sources of information handle the input to and from the cell. The main advantage of LSTM is that it can learn to establish implicit transient dependencies. Each LSTM unit stores long-term or short-term data (hence its name) without explicit use of initial operations in the passing section. It is important to note that the state of each cell is significantly increased by the result of the neglected terms, shifting somewhere between 0 and 1 [8 Overall, the above a transferred to the LSTM cell causes the initialization of the cell and the active position of the heap. In this way, information from the state in front of the cell does not increase or decrease significantly with each time step or phase, nor can it pass through the cell unchanged, and the charge does not change at within a reasonable time frame to achieve the best quality. This allows LSTM to solve the problem of missing bias. Because values stored in memory cells do not change frequently, biases do not disappear when searching for causal relationships. Markets like NSE and BSE have been established as Indian element exchanges. To cut us off.



Alogorithm	Accuracy(%)
Logistic Regression	56.09
Support Vector Machine	51.32
LSTM	85.50
Random Forest Classifier	81.24
Linear Regression	53.3

X. CONCLUSION

In summary, for this purpose, our study uses various models such as long short-term memory (LSTM), support vector machine (SVM), random forest classifier (RFC), stock linear regression and logistic regression. We conducted a thorough investigation using Price prediction. The results demonstrated the different strengths of each model in capturing different aspects of the complex dynamics of financial markets. LSTM is a type of recurrent neural network (RNN) that has demonstrated commendable performance in detecting time-dependent and complex patterns in time-series data. Support vector machines demonstrated robustness in handling nonlinear relationships, and random forest classifiers demonstrated strong predictive capabilities through ensemble learning. Despite their simplicity, linear and logistic regression provided valuable insights into the linear relationships and binary classification aspects of stock price movements. Nevertheless, finding a generally good model remains a challenge, as each algorithm has its own strengths and limitations. Ongoing challenges include the nonlinear nature of market dynamics, the impact of unforeseen events, and the ever-evolving nature of market sentiment that impacts all models. Therefore, a practical approach should exploit the strengths of different models through ensemble or hybrid approaches, thereby creating a more robust and accurate predictive framework. Ultimately, choosing the most appropriate model depends on the specific characteristics of your dataset and the goals of your prediction task. Our results highlight the importance of diverse and adaptive approaches to stock price forecasting, improving existing models and exploring new techniques to In precis, for this motive, our look at uses various fashions inclusive of long brief-time period reminiscence (LSTM), aid vector machine (SVM), random forest classifier (RFC), stock linear regression and logistic regression. We performed a radical research the usage of Price prediction. The outcomes proven the exclusive strengths of each version in capturing unique factors of the complicated dynamics of economic markets. LSTM is a type of recurrent neural network (RNN) that has confirmed commendable performance in detecting time-based and complicated styles in time-series records. Support vector machines proven robustness in handling nonlinear relationships, and random woodland classifiers validated robust predictive abilities through ensemble learning. Despite their simplicity, linear and logistic regression furnished precious insights into the linear relationships and binary class factors of stock rate actions. Nevertheless, locating a typically right version remains a assignment, as each set of rules has its very own strengths and barriers. Ongoing challenges consist of the nonlinear nature of market dynamics, the effect of unforeseen activities, and the ever-evolving nature of marketplace sentiment that affects all models. Therefore, a sensible technique should exploit the strengths of different fashions via ensemble or hybrid procedures, thereby growing an improved and correct predictive framework. Ultimately, choosing the maximum appropriate version relies upon on the specific characteristics of your dataset and the dreams of your prediction mission. Our consequences highlight the importance of diverse and adaptive strategies to inventory rate forecasting, improving present models and exploring new strategies to enhance forecast accuracy within the dynamic context of economic markets [9].

REFERENCE

[1] Polamuri Subba Rao, K. Srinivas, and A. Krishna Mohan. Department of CSE, VR Siddhartha Engineering College, Vijayawada, AP. A Survey on Stock Market Prediction Using Machine Learning Techniques. (January 2021)

[2] Pramod B S, Mallikarjuna Shastry P. M.. M tech [pt] 6th Semester in CSE, REVA University, Bengaluru. Stock Price Prediction Using LSTM. (May-June 2020)

[3] Jaydip Sen Department of Analytics and Information Technology Praxis Business School Kolkata, INDIA. Stock Price Prediction Using Machine Learning and Deep Learning Frameworks. (December 2018)

[4] Seyda Kalyoncu, Akhtar Jamil, Enes Karatas, Jawad Rasheed, Chawki Djeddi. Department of Computer Engineering, Istanbul Sabahattin Zaim University, Istanbul, TURKEY. Department of Mathematics and Computer Science, Larbi Tebessi University, Tebessa, Algeria. Stock Market Value Prediction using Deep Learning. (May 2020)

[5] R. Sathishkumar, R. Girivarman, S. Parameswaran, V. Sriram, Assistant Professor, Department of Computer Science & Engineering, Manakula Vinayagar Institute Of Technology, Puducherry, Student, Department of Computer Science & Engineering, Manakula Vinayagar Institute Of Technology, Puducherry. Stock Price Prediction Using Deep Learning and Sentimental Analysis. (October 2020)

[6] Yogita Deshmukh, Deepmala Saratkar, Harshal Hiratkar, Sudhanshu Dhopte, Swapnil Patankar, Triveni Jambhulkar, Yash Tiwari Assistant Professor, Computer Technology, RGCER, Nagpur, Maharashtra, India. Stock Market Prediction Using Machine Learning. (January 2019)

[7] V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India. Stock Market Prediction Using Machine Learning. (October 2018)

[8] Zahra Fathali, Zahra Kodia & Lamjed Ben Said (2022) Stock Market Prediction Applying Machine Learning Techniques, Applied Artificial Intelligence. (September 2022)

[9] Jaydip Sen, Sidra Mehtab, Gourab Nath Department of Data Science Praxis Business School Kolkata, INDIA. Stock Price Prediction Using Deep Learning Models. (November 2020).