Forecasting Stock Prices Using Time-Series Analysis, Regression Learning, and Deep Learning Models

KEKKERENI GANESH GOUD
Computer Science and Engineering
ganeshkekkireni7731@gmail.com

BANDI SHIVARAMA DEEKSHITH
Computer Science and Engineering
shivrambandi@gmail.com

***Abstract*— For an extended period, scholars have been crafting a dependable and precise forecasting model for predicting stock prices. Predictive models may meticulously and accurately anticipate future stock prices if they are properly created and improved, according to the literature. This paper presents a number of learning-based, econometric, and time series models for predicting stock prices. Here, the data from WIPRO, SBI, and APOLLO PHARMA from January 2000 to December 2021 was utilized to train and test the models in order to determine that algorithms worked in what industry. This study includes two Regression models (Random Forest and MARS), two deep learning-based models (basic RNN and LSTM), one econometric model (ARIMA), and one-time series model (Holt-Winters Exponential Smoothing). It has been demonstrated that LSTM is the best deep learning model and that MARS is the greatest machine learning model. But overall, MARS has shown to be the greatest-performing model in sales forecasting for all sectors: Information technology (using values from WIPRO), Finance (using values from SBI), and Health (using data from APOLLO PHARMA).**

Keywords*— *Time Series Analysis, Forecast stock price, Recurrent Neural Networks, LSTM, Random Forest, MARS.

I. Introduction

For a very long time, experts have been fascinated with predicting future stock prices. Numerous research contends that precise stock price estimation is unattainable and refutes the efficient market theory. According to several claims made in the study, if finding models are properly developed and then optimized, they are used to find future prices of stock with reasonable accuracy and consistency. Furthermore, it has been discovered that the variables selected, the methods used, and the manner in which the model was refined all have an impact on how accurate a predictive model is. Utilizing the time series breakdown, scientists proposed a method for predicting stock prices in this area. Machine learning and deep learning are becoming more and more popular in addition to predicting stock prices and observing the trends in their movements. Additionally, a Time-Series Analysis-based technique was applied to do a study of the banking, real estate, high-cap, medium-cap, and other Indian industries. Scientists have also found a good and robust-founded prediction model for Google's stock price forecasting utilizing an RNN-based neural network approach. The current study uses a mix of time series analysis(TS), and learning-based models to anticipate the future values of three key stocks listed on the National Stock Exchange (NSE) of India. We looked at the stocks of APOLLO PHARMA, SBI, and WIPRO. The study focuses on three industries: IT, finance, and health. For the finance industry, SBI's stock price values was taken, WIPRO's stock price values were utilized for the information technology (IT) sector study, and APOLLO PHARMA's stock pricing data was taken into account for the health industry. There have also been proposals in the literature for using Regression models learning, including neural networks learning models, as well as language processing on social media news items.

Three significant stocks of the National Stock Exchange (NSE) of India are the subject of the current study, which uses a combination of time series analysis (TS), machine learning-based models to forecast their next values. We looked at the stocks of Apollo Pharma, SBI, and Wipro. The IT, financial services, and health industries are the three areas of focus of the research. ICICI's stock price data was selected for the banking industry, Infosys' stock price data was used for the IT sector research, and APOLLO PHARMA's stock price data was taken into consideration for the health sector. Data for each of the aforementioned businesses was gathered from January 2004 to December 2019. The main study of our research is as follows. Initially, we combined Regression learning and neural learning techniques to construct four models that could predict stock prices with a very high degree of precision. After that, we used those models in three significant industries. By examining the stock values from the preceding seven days, we can use the LSTM model to forecast the value of the shares for the next eight days. In the end, this enables us to identify which model, out of each one that has been employed, is most successful in a certain industry. The best-performing models are identified through the smallest proportion of the RMSE to the average closing stock price.

DATASETS

We require a dataset containing the following data in order to train using machine learning (RF, MARS) as well as deep learning (RNN, LSTM) for stock price prediction. Historical Data on Stock Prices: It is necessary to build predictive models utilizing historical stock prices, including high, low, open, and close data. This data is being used to look for patterns and trends. Trading Volume Information about a stock's volume of trading might reveal details about the market's interest and liquidity. It is often used in conjunction with pricing data when generating forecasts.

Characteristics for training a prediction model				
Date	Open Value	High Value	Low	Close

Table 1: attributes for the data set.

Date: This column typically contains the day when the stock price data was recorded. Open Price: The opening price of the stock on the designated day. High Price: The highest price the stock reached in the course of the trading session. Low Price: The price at which the stock fell throughout the trading day. Close Price: The closing price of the stock on the designated day. Volume: The total quantity of shares or contracts traded in a given day. It represents the trade activity during the day.

II. Problem Statement

Our study aims to develop a practical model for forecasting stock prices in the banking, healthcare, and information technology sectors. Thus, from the beginning of 2000 to December 2021, Kaggle provided about 20 years' worth of data for three companies in the IT, banking, and health sectors: WIPRO, SBI, and APOLLO PHARMA. A combination of machine learning, deep learning, econometric, and time series models was used to predict the stock prices of the three companies listed above. We predict that the MARS algorithm is found as the good accuracy in all the employed machine learning models due to its ability to find important characteristics in a data set and create an algorithm that combines many linear functions. Owing to its ability to extract a wide range of attributes from data, the model developed by LSTM is also expected to yield exceptional outcomes.

III. Related Work

Three categories are included in the paper to categorize the current state of stock price forecasting research. The start section for the models uses the Random Forest Regression technique to handle multivariate data. This method is effective in generating short-term predictions, however, it has certain drawbacks, consisting of the adjustment of seasonal indexes, the selection of opening values, and the selection of best features. The primary drawback of the second sort of work, which also used the ARIMA model, is that before using ARIMA, the sequence must be made stationary. Moreover, financial time series exhibit a rise in instability, that isn't covered by the ARIMA assumption, although the ARIMA model implies constant variance The third class of models uses machine learning techniques like MARS and the random forest model. Random Forest regression is not very good at forecasting when there is volatility in the data. Nevertheless, MARS enhances the output of regression issues by eliminating predictor variables that have no effect on the model. The fourth category in our analysis consists of models based on deep learning like RNN and LSTM, which are able to estimate stock values extremely precisely because they can learn nonlinear patterns from historical data. As a consequence, uncertainty in the financial time series may be easily addressed.

IV. METHODOLOGY

The techniques we employ in this study to forecast the value of stocks are learning-based techniques. The algorithms RNN and LSTM are used in the deep learning approach, whereas Mars and the random forest regressor are used in machine learning. To test these models, all of the previously mentioned algorithms were used in three different industries for the stocks value of three firms: APOLLO PHARMA, SBI, and WIPRO. The Kaggle website provided the stock values for all three companies from January 2000 to the end of 2021. The date, open, high, low, and close are the factors that make up the stock values. The closing date of the stock price was used for the goal parameter for all models.

A. Random Forest Regression

Random Forest Prediction is an ensemble technique that estimates the output of a random forest framework by averaging all of the forecasts for each decision tree. It applies the aggregation and bootstrap bagging concepts. The term "bootstrap" describes the process of selecting random samples with replacements from a dataset. Aggregation is the process of combining all forecasts to generate the ultimate result. Reducing overfitting in models is aided by bagging. Because Random Forest is a well-liked ensemble learning approach that produces robust and high-accuracy outcomes, it has been used extensively in stock price prediction. Its achievement may be ascribed to a number of crucial elements that improve performance in this field. Combination of Decision Trees: A model of ensembles made up of many decision trees is called a Random Forest. A random portion of the data set is used for training each tree, and the outcomes are averaged or blended by majority vote. Effective stock price prediction depends on the model's capacity to reduce overfitting and boost generalization, both of which are enhanced by the ensemble technique. Random Forest offers a method for determining the significance of a characteristic. This implies that it can determine which financial elements or indications have the most bearing on accurate stock price forecasts. Refinement of the model and feature selection benefit greatly from this knowledge. Managing Non-Linearity: It is well known that the behavior of stock prices is complicated and non-linear. Random Forest is an excellent choice for modeling the complex patterns and trends that affect stock values since it can successfully capture these non-linear interactions. Robustness against Outliers: Outliers are not unusual in financial data, which can be noisy at times. Because a Random Forest depends on the predictions of several trees, it is naturally resistant to outliers and lessens the influence of each of the noisy data points. Cross-validation: The cross-validation techniques are frequently used to evaluate a model's correctness and make necessary adjustments. Because Random Forest can handle numerous subsets of data, it does a good job at cross-validation, making sure the algorithm generalizes to new data successfully.

For the purpose of creating a model, multivariate values were run upon open value, high value, low value, close value, adj-close, and volume variables. A training dataset consisting of 5708 records was obtained from SBI, APOLLO PHARMA, and WIPRO, while a test dataset consisting of 241 records was obtained from 2021. To place all of the characteristics in the 0–1 range, MinMax scaling was applied. The X-test was used to predict the near-column values, and the random forest algorithm was constructed on the X and Y trains. In relation to the original scale, the figures were scaled inversely. The train of close mean and the ratio of RMSE score to the test of close mean were calculated while evaluating the y-test and y-pred.

B. MARS

The multivariate adaptive regression spline modifies the volatility of the data set. The input variables are split up into several step functions using the MARS approach. We call them fundamental functions. Data cut points referred to as knots are used to assess these functions. For every value, the model finds for an array of numbers and chooses a step function that yields the smallest missed data. To generate a reliable non-linear prediction algorithm, a pivot function is subsequently used at every knot, and the procedure is then repeated. Training datasets are going to fit well if there are more knots, but this might lead to overfitting. When a model is overfitted, test data findings become less reliable, necessitating cross-validation model pruning to determine the ideal knot count. The idea behind the MARS (Multivariate Adaptive Regression Spline) operational concept is to integrate step functions to match at every value to create a complex model. The forward passing phase is when the knots need to be located. During the backward iteration phase, the model wants to reduce or remove words that are low contributes in form to reduce noise, and overfitting.

For model-building purposes, a multivariate analysis was conducted on open, high, low, close, adj-close, and volume. A training dataset consisting of 5708 records was obtained from SBI, APOLLO PHARMA, and WIPRO, while a test dataset consisting of 241 records was obtained from 2021. To place all of the attributes in the 0–1 range, minmax scaling was applied. The earth() function was selected from the py-earth package in order to execute this MARS algorithm. The X-test was used to predict the near-column values, and the MARS model was developed using the X-train and y-train. In relation to the original scale, the numbers were scaled inversely. The ratio of the root-mean score and test value for the close was found by contrasting the y-test to the y-pred. The third class of models uses machine learning techniques like MARS and random forest model. Random Forest regression is not very good at predicting when there's volatility in the data. Nevertheless, MARS enhances the output of regression problems by automatically removing variables that predict that have no effect on the model.

C. RNN

Recurrent neural networks, also referred to as RNNs, are becoming an indispensable instrument in the deep learning space because of their special ability to evaluate sequential input. RNNs have a unique design that enables them to retain a memory of previous input, which sets them apart from standard feedforward neural networks. This feature makes them highly suitable for jobs requiring time-series data, natural language processing, and other sequential applications. The main novelty of RNNs is that they can handle input in a sequence and continuous context since they may include feedback loops. RNNs are able to capture complex patterns, relationships, and temporal details within the data because of their recurring nature, which is advantageous for a variety of real-world applications. RNNs are quite flexible, but they have drawbacks. For example, disappearing and expanding gradient difficulties. As a result, more sophisticated versions such as LSTM (Long Short-Term Memory) and Gated Recurrent Unit, or GRU, networks were created. These developments have greatly improved RNNs' ability to model and forecast data over time, making them a priceless tool in the toolbox of ML methods.

Recurrent neural networks, or RNNs, have advanced significantly in many different applications, but what makes them special is their ability to process dynamic, variable-length sequences. RNNs are unique in that they can adjust to inputs of varying durations, which makes them appropriate for jobs where sequences might vary greatly in length, such as language models and speech recognition. The network's recurrent connections let it process data sequentially, component by part while remembering previous inputs, which contributes to its flexibility. RNNs are particularly good at modeling long-term context and short-term interdependence within a sequence. This makes them useful for applications like sentiment analysis, where determining a sentence's sentiment may depend on its preceding words or phrases. Moreover, a generative model that powers advances like text creation and music composition is centered upon RNNs that are They have been extremely helpful in the development of virtual assistants and chatbots that can have dynamic, context-aware discussions and provide a distinctive, interactive user experience. Because of its dynamic nature and capacity to recognize intricate temporal patterns, RNNs are an effective tool for producing and comprehending sequential data, which improves applications in a variety of fields.

Managing Sequential Data: RNNs were excellent at model sequential connections since trading data is by its very nature sequential. They are able to identify intricate trends, patterns, and temporal correlations that affect stock prices. RNNs are able to anticipate future prices by analyzing historical stock prices and extracting relevant information from past data. Feature Engineering: A key component of stock price prediction is feature engineering. Lagged pricing, technical signs, and external influences (such as news emotions or economic data) are just a few of the many elements that RNNs may contain. By carefully choosing and designing these elements, RNN models' accuracy may be greatly increased. Standardization and Enlargement: RNNs operate with constant input data when proper data preparation, such as scaling and normalization, is carried out. For the system to train efficiently and generate precise predictions, this is necessary. Design Architecture: Performance may be significantly impacted by the RNN's design, which includes the number of layers, the units for every layer, and the selection of activation functions. The secret to getting the best results is to experiment with alternative designs and hyperparameters.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

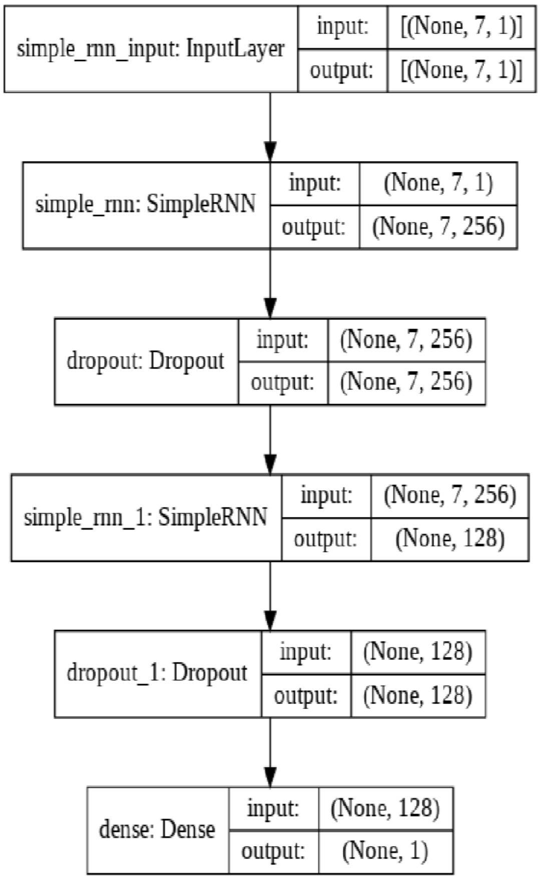


Fig. 1. The RNN algorithm for forecast stock prices

As a performance metric for assessing a predictive model's accuracy, the Root Mean Square Error (RMSE) is computed in two steps: first, the squared errors are computed by taking the difference between the actual target values (y_test) and the predicted values (y_pred), then squaring the differences. These squared errors quantify the amount of error that the model makes in each prediction for each data point. Next, the mean of the squared errors is computed, which gives an overall measure of the average error throughout the dataset. This is done by adding up all of the squared errors and dividing the resultant amount by the number of data points. The squared root of the mean square error is calculated to get the final RMSE. A popular statistic for evaluating the precision of prediction models, particularly in regression tasks, is the root mean square error (RMSE). It is given in the same units as the target variable and gives an indication of the typical size of prediction mistakes. The computed root mean square error (RMSE) is around 8.07. In the same units as the target variable, this number indicates the model's normal prediction error. Greater RMSE values imply that the algorithm's predictions are less accurate, whereas lower RMSE values show superior accuracy in forecasting. As a result, the RMSE value is a crucial assessment indicator for determining how well the predictive model performs.

D. LSTM

The specialized recurrent neural network, or RNN, architecture known as Long Short-Term Memory (LSTM) has attracted a lot of interest because of its remarkable capacity to model and predict sequential input. Traditional RNNs suffer from disappearing and expanding gradient issues, which LSTMs are specifically intended to handle. The memory cell of an LSTM, which stores and refreshes data over time to enable it to recognize long-term relationships in sequences, is fundamental to the device's efficacy. Three gates are included in this memory cell: an input gate, an output gate, and a forget gate. As LSTMs analyze sequential input, these walls protect the passage of data, allowing them to store or reject data selectively. The forget gate chooses which data should be deleted, the output gate decides which data should be utilized to produce predictions, and the input gate manages the addition of new data to the cell state. Because of their gated architecture, long short-term memory (LSTM) systems are especially well-suited for a variety of uses, including speech recognition, natural language processing, time series analysis, and, most notably, stock price prediction. LSTMs are particularly effective at capturing complex patterns and relationships in sequential data. The long-range dependability of LSTM and its resilience to gradient-related problems have made it a powerful tool in a deep learning toolkit, facilitating the creation of more reliable and accurate models for sequential data processing. Because of its distinct architectural features, LSTM (Long Short-Term Memory) algorithms have been shown to be a potent and useful tool for stock price prediction. Retaining the intricate temporal correlations and patterns included in finance time series data is a major obstacle in stock price prediction. Because they solve the disappearing and expanding gradient issues that frequently plague conventional RNNs, LSTMs perform exceptionally well in this area. Specialized memory cells with gating mechanisms are a feature of the LSTM design that enables the model to store, retrieve, and forget data over long sequences. This feature is essential for simulating stock prices since it allows the network to filter out noise and capture both long- and short-term trends and volatility.

To provide a complete picture of the variables affecting stock prices, LSTMs can also include other pertinent aspects including external market data, technical indicators, and lagged stock prices. This characteristic engineering flexibility adds to the correctness of the model. Moreover, LSTMs work well with noisy and irregular data, which is typical in financial markets. Because they can learn from past data, they are strong and adaptable for stock price prediction, allowing them to adjust to shifting market circumstances and unforeseen occurrences. Because of its capacity to handle sequential data and simulate complex relationships in stock prices, long short-term memory banks (LSTMs) are a top option for creating precise and dependable financial market prediction models. It's crucial to remember that predicting stock prices is still a difficult undertaking because a wide range of factors, such as market mood, unforeseen occurrences, and economic conditions, can affect it. LSTMs are a useful tool, but in order to make well-informed investment decisions, they should be utilized in concert with other analytical methods and risk management approaches.

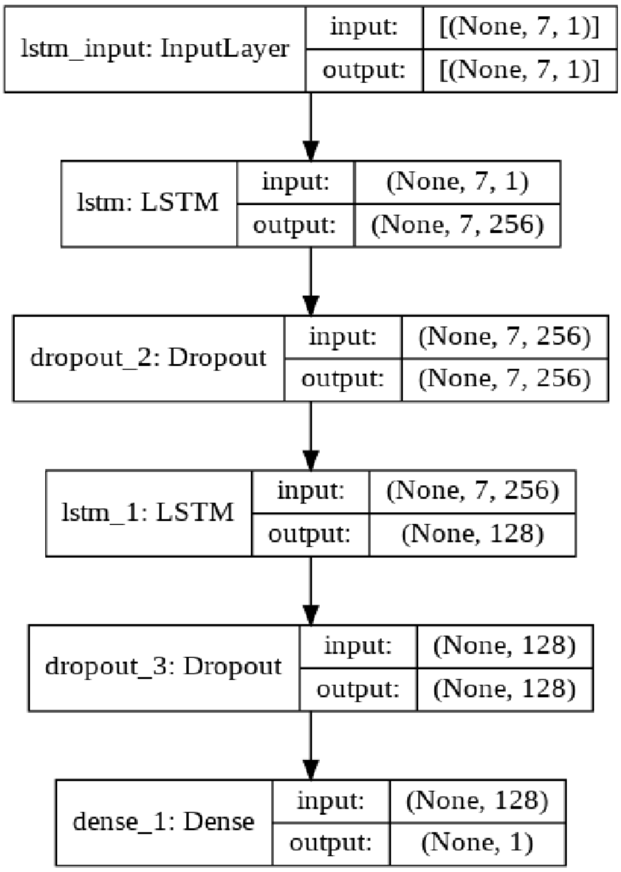


Fig. 2. The LSTM algorithm to forecast stock price

There were 241 entries in the test data set (2021) compared to 5708 entries in the train set (2000 to 2020). We reduced the training data to 4700 days using a seven-day rolling window, and the testing data to 547 days. Using stacked LSTM and simple RNN, the prediction was done on the shortened data. It's critical to stress that predicting stock prices is difficult by nature since it depends on a variety of outside variables, market mood, and unpredictable occurrences. Although long-term moving averages (LSTMs) may be an effective tool for traders and financial analysts, they should be utilized cautiously and in conjunction with other methods, risk management techniques, and due diligence when making investment decisions.

V. Experimental Results

The final result shows how well four models that were created to predict the stock prices of three significant industries IT, banking, and healthcare performed. The effectiveness of the models was evaluated using an evaluation measure that computes the proportion of the root mean and test mean of nearby readings for the stock price. The average of neighboring train data values was mean for the train set. The average of the test data's near values was the mean for the test set. The following is a list of each model's performance outcomes used in each industry.

A. IT Sector

The closing value of WIPRO's shares was used in the technology sector. A total of four algorithms were used for finding the prices of stock. Every model's result is displayed, along with a table including the evaluation measure for each model.

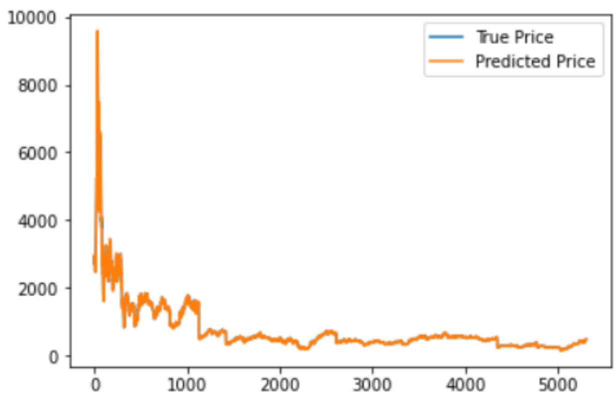


Fig. 3 (a) True and Predicted Prices of stock for WIPRO with Random Regression algorithm.

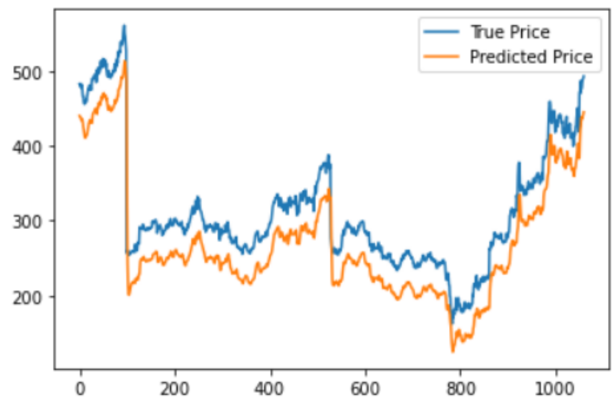


Fig. 4 (a) True and Predicted Prices of stock for WIPRO with Recurrent neural network.

B. Finance Sector

In the financial services industry, our analysis took into account the closing value of SBI's stock price. A total of four algorithms were used to predict the prices of stock. Every model's result is displayed, along with a table including the evaluation measure for each model.

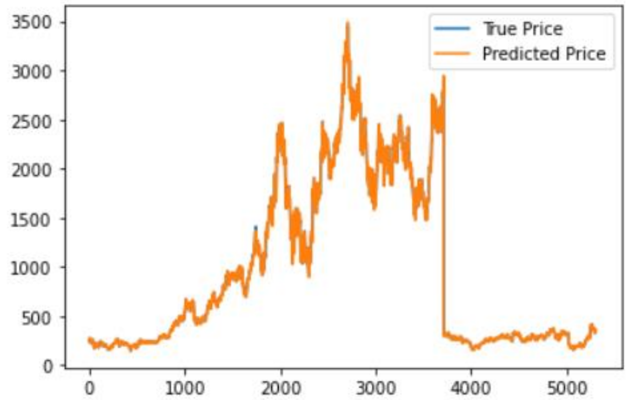
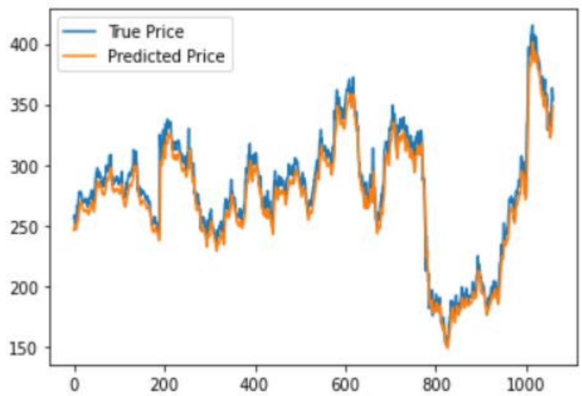


Fig. 5(b) True and Predicted Prices of stock for SBI with Random Regression algorithm.



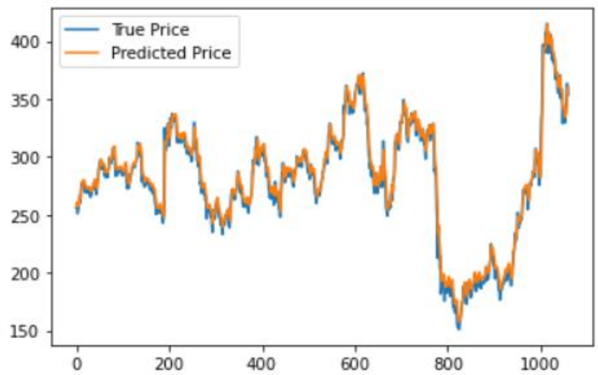


Fig. 5 (d) True and Predicted Prices of Stock for APOLLO pharma with Recurrent neural network.

Fig. 6(b) True and Predicted Prices of stock for SBI with\ Recurrent neural network

C. HealthCare Sector

Our analysis took into account the near value of Apollo Pharma's stock price in the medical industry. A total of six models were used to predict the stock price. Every model's output is displayed, along with a table including the evaluation measure for each model.

Fig. 6(b) True and Predicted Prices of stock for Apollo Pharma with Random Forest

It is clear that the Random Forest algorithm has produced varied Root Mean Square Error (RMSE) values for different industries when it comes to predictive models and performance evaluation. Particularly in the field of financial analysis, root mean square error (RMSE) is an essential indicator for evaluating the precision and efficacy of prediction models. These RMSE values provide insight into the model's prediction power and industry applicability. The Random Forest model produced an RMSE of around 11.69 for the IT industry. This number represents the average size of prediction mistakes for data from the IT industry. The larger RMSE suggests that stock price prediction in the IT industry may be more difficult, maybe as a result of the industry's higher volatility and degree of unpredictability. On the other hand, the Random Forest model shows a substantially lower RMSE of around 4.48 in the healthcare industry. The model's increased accuracy in forecasting the price of stocks for the healthcare industry is indicated by the decreased RMSE. This may be explained by the industry's tendency toward stability and less volatility, which enables the model to provide more accurate forecasts. The Random Forest model yields an RMSE of about 5.45 for the banking industry. The model's prediction performance appears to be in the middle of the IT and healthcare industries, based on this intermediate RMSE score. The mild volatility of banking stocks suggests that the model is useful in this industry, as seen by its ability to produce forecasts with this degree of accuracy. The RNN model yields an RMSE of around 43.68 for the IT industry. This comparatively large RMSE indicates that RNNs may have substantial difficulties in forecasting stock values in the information technology sector. Accurate projections become more difficult in this industry because of its high volatility, quick technology improvements, and vulnerability to market emotion. On the other hand, RNNs have a significantly lower RMSE of around 8.07 when used in the healthcare industry. The RMSE of RNN models is around 11.25. The predictive accuracy of RNNs lies between the IT and healthcare industries, according to this intermediate RMSE value. Since the banking industry usually experiences modest levels of volatility, the RNN model's usefulness in this industry is demonstrated by its capacity to produce predictions with this degree of precision.

D. The Comparison of each algorithm's effectiveness across the three sectors

By finding the proportion of RMSE to the average of the close data of the prices of stock, a comparative analysis using the results of every single model across every single industry has been examined and is shown in a tabular style.

Stock Sector	Algorithms	Root Mean Square Error/ Average of Close value of test set
Information Technology Sector	Random Regression	0.034
	MARS	0.0068
	RNN	0.0114
	LSTM	0.012
Finance Sector	Random Regression	0.0714
	MARS	0.077
	RNN	0.024
	LSTM	0.0113
HealthCare Sector	Random Regression	0.007
	MARS	0.021
	RNN	0.020
	Ls	0.011

Random forest, RNN, and LSTM models are applied in many sectors. Regression and classification may be accomplished with the use of efficient machine-learning models like LSTM, RNN, and Random Forest. But each of them has a unique set of advantages and disadvantages. Random forest models are renowned for their accuracy and ability to withstand overfitting. They are quite easy to learn and understand. But when it comes to training on large datasets, they may be slow. RNN models are useful for tasks like machine translation and natural language processing, for example. Long-term data dependencies are another thing they may learn. Conversely, RNNs can be challenging for training and are susceptible to overfitting. The information technology, financial services, and medical industries exhibit the following performance patterns: Random Forest: Does well in the banking sector(SBI), but not so well in the IT sector(WIPRO). RNN Does well in the medical field but not so well in the IT field. LSTM: Competitive in all three, but provides exceptional outcomes in the IT industry. This demonstrates that LSTM models are the most versatile and effective machine learning models for use in a wide range of sectors. The capabilities of LSTM (Long Short-Term Memory) networks a flexible and efficient.

LSTM is a great option for precise forecasts since it is adept at predicting stock prices by recognizing intricate temporal relationships and adjusting to shifting market conditions.

VI. Conclusion

The paper used many approaches to forecast stock prices. In order to predict the stock price of three different sectors, this study employed a Random Regression model, MARS models, and RNN, and LSTM learning-based algorithms. Between January 2000 and the end of 2020, a set of training data was used to develop, train, and improve the models. The 2021 test dataset was used to evaluate the trained model. The results indicate that MARS is the most dependable machine learning model and LSTM is the finest deep learning model, despite the fact that all of the models utilized attained exceptional levels of accuracy. Its capacity for feature selection and application of variability from a MARS dataset contributed to its extraordinary performance. Conversely, LSTM yielded a positive result because of its ability to build a model that can handle complex sequential data without experiencing problems with diminishing and growing gradients. However for the consists to reduce the greatest values of attributes and then subsample them, as well as its rapid implementation speed, which will lead to outstanding finding performance, we aim to integrate CNN-based prediction of stock prices in our future study.

References

[1] J. Sen and T. Datta Chaudhuri, “An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice - a comparative study of the Indian consumer durable and smallcap sector”. *Journ.of Eco. Lib.* vol. 3, no. 2, pp. 303 - 326, 2016.

[2] J. Sen and T. Datta Chaudhuri, “Decomposition of time series data of stock markets and its implications for prediction - An application for the Indian auto sector,” *Proc. of the 2nd Nat. Conf. on Adv. in Bus. Res. and Mgmt Pract*, Kolkata, India, pp. 15-28, 2016.

[3] J. Sen and T. Datta Chaudhuri, “An investigation of the structural characteristics of the Indian IT sector and the capital goods sector – An application of the R programming language in time series decomposition and forecasting”, *Jour. of Ins. and Fin. Mgmt*, vol. 1, no. 4, pp. 68-132, June 2016.

[4] J. Sen and T. Datta Chaudhuri, "A time series analysis-based forecasting framework for the Indian healthcare sector", *Jour. of Ins. and Fin. Mgmt,* vol. 3, no. 1, pp. 66-94, 2017.

[5] I. Parmar et al., "Stock Market Prediction Using Machine Learning," *Proc. of the 1st Int. Conf. on Secure Cyber Comp. and Comm.,* Jalandhar, India, 2019.

[6] J. Shen and M. O Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system", *Journ. of Big Data*, vol. 1, 2020.

[7] J. Sen, "Stock price prediction using machine learning and deep learning frameworks", *Proc. of the 5th Int. Conf. of Bus. Analytics and Intl, Bangalore* Dec 11-13, 2017, Bangalore, India.

[8] J. Sen. and T. Datta Chaudhuri, "A predictive analysis of the Indian FMCG sector using time series decomposition-based approach", *Journ. of Eco. Lib.,* vol. 4, no. 2, pp. 206-226, June 2017.

[9] J. Sen and T. Datta Chaudhuri, "Understanding the sectors of Indian economy for portfolio choice", *Int. Journ. of Bus. Forecast. and Mktg. Intel.,* vol. 2, no. 2, pp. 178-222, 2018.

[10] J. Sen, "A robust analysis and forecasting framework for the Indian mid-cap sector using time series decomposition", *Journ. of Ins. and Fin. Mgmt*, vol. 3, no. 4, pp. 1- 32, 2017.

[11] K. Ullah, M. Qasim, "Google stock prices prediction using deep learning", *Proc. of the IEEE 10th Int. Conf. on Sys Engg. and Tech.*, Nov 9, 2020, pp. 108-113, Shah Alam, Malaysia.

[12] S. Mehtab and J. Sen, "Stock price prediction using convolutional neural network on a multivariate time series", *Proc. of the 3rd Nat. Conf. on Mach. Lrng. and Art. Intel*, February 1-2, New Delhi, India, 2020.

[13] J. Wu, Z. Li, N. Herencsar, B. Vo, and J. Lin, "A graph-based CNNLSTM Stock price prediction algorithm with leading indicators," *Multimedia Systems,* January 2021(Accepted).

[14] S. Mehtab and J. Sen, "A time series analysis-based stock price prediction using machine learning and deep learning models", *Int. Journ. of Bus. Forecast. and Mktg Intel. (IJBFMI)*, vol. 272, no. 4, pp. 272-335, 2020, Inderscience.

[15] S. Mehtab and J. Sen, "Stock price prediction using CNN and LSTMbased deep learning models", *Proc. of IEEE Int. Conf. on Decs. Sc. and Appln.*, Nov 8-9, 2020, Sakheer, Bahrain, pp. 3447-453.

[16] S. Mehtab, J. Sen, and A. Dutta, "Stock price prediction using machine learning and LSTM-based deep learning models". *Machine Learning and Metaheuristics Algorithms, and Applications*, pp. 88-106, Springer, Singapore.

[17] S. Mehtab, J. Sen, and S. Dasgupta, "Robust analysis of stock price time series using CNN and LSTM-based deep learning models", *Proc. of Int. Conf. on Electronics, Comm., and Aerosp. Tech.*, Nov 5-7, 2020, Coimbatore, India, pp. 1481-1486.

[18] J. Sen, A. Dutta, and S. Mehtab, "Profitability analysis in stock investment using an LSTM-based deep learning model", *Proc. of 2nd IEEE Int. Conf. for Emerg. Tech.*, May 21-23, 2021, pp. 1-9, Belagavi, India.

[19] J. Sen and S. Mehtab, "Accurate stock price forecasting using robust and optimized deep learning models", *Proc. of the IEEE Int. Conf. of Intel. Tech.,* June 25-27, 2021, Hubbali, India. (Accepted for publication)

[20] S. Mishra, "The quantile regression approach to analysis of dynamic interaction between exchange rate and stock returns in emerging markets: Case of BRIC nations", *IUP Journ. of Fin. Risk Mgmt*, vol. 13, no. 1, pp. 7 – 27, 2016.

[21] Y. Ning, L. C. Wah, and L. Erdan, "Stock price prediction based on error correction model and Granger causality text", *Cluster Comptg*, vol. 22, pp. 4849-4858, 2019.

[22] L. D. S. Pinheiro and M. Dras, "Stock market prediction with deep learning: A character-based neural language model for event-based trading", *Proc. of the Aust. Lang. Tech. Assoc. Workshop*, Brisbane, Australia, pp. 6 -15, 2017.

[23] S. Mehtab and J. Sen, "A robust predictive model for stock price prediction using deep learning and natural language processing", *Proc. of the Int. Conf. on Bus. Analytics and Intell.*, Dec 5-7, 2019, Bangalore, India.

[24] M. H. Rahman, M. M. Hossain, U. Salma, M. T. F. Khan, "Revenue forecasting using Holt-Wniters exponential smoothing", *Research & Reviews: Journ. of Stats.,* vol 5, no. 3, pp.19-25, 2016.

[25] T. Vantuch and I. Zelinka*, "*Evolutionary based ARIMA models for stock price forecasting*", Proc. of Interdisci. Symp. on Compl. Sys.,* pp, 239-247, 2014.

[26] S. R. Polamuri, K.Srinivas, A. K. Mohan, "Stock market prices prediction using random forest and extra tree regression", *International Journ. of Recnt. Tech. and Engg.*, vol. 8, no. 3, pp. 1224-1228, September 2019.

[27] C. Lu, C. Chang, C. Chen, C. Chiu, "Stock index prediction: A comparison of MARS, BPN, and SVR in an emerging market", *Proc. of IEEE Conf. on Indust. Engg. and Engg.Mgmt,* Dec 8-11, 2009, Hong Kong, China, pp. 2343-2347.

[28] A. Moghar, M. Hamiche," Stock market prediction using LSTM recurrent neural network", *Procedia Comp. Sc.*, vol. 170, pp. 11681173, 2020.

[29] S. Selvin, R. Vinaykumar, E. A Gopalkrishnan, V. K. Menon, K. P. Soma, "Stock price prediction using LSTM, RNN, and CNN-sliding window model", *Proc. of the Int. Conf. on Adv. in Comptg.*, Comm., and Inform., Sep 13-16, 2017, Udupi, India, pp. 1643-1647. [30] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root", *Journ. of Econometrics*, vol. 54, nos. 1-3, pp. 159-178, 1992.

[30] J. Sen and T. Datta Chaudhuri, "Understanding the sectors of Indian economy for portfolio choice", *Int. Journ. of Bus. Forecast. and Mktg. Intel.,* vol. 2, no. 2, pp. 178-222, 2018.