

Stock Price Prediction by using Machine Learning Techniques: a Study of TCS Ltd

1st Yogesh Kumar Jakhar

Department of Computer Application
University of Engineering &
Management
Jaipur, India
yogeshjakhar@gmail.com

2nd Pawan Sharma

School of Management
University of Engineering &
Management
Jaipur, India
sharma.sirsa@gmail.com

3rd Bilal Ahmed

School of Computer Application
Lovely Professional University
Phagwara, India
bilal.ahmed.jjn@gmail.com

Abstract—Prediction of stock prices is challenging, but now stock price prediction is becoming very popular among researchers. Exact prediction of the stock market is not feasible, but the present study is an attempt to understand stock price prediction with special attention to TCS Ltd. The key objective of the present study is to forecast the high price of TCS stock on a daily basis. To achieve the objective successfully, this study comprises the use of historical time-based data for 15 years, from 2009 to 2023. (12 years of data for training purposes and 3 years of data for testing purposes are significantly categorized). Day-wise high prices have been significantly predicted by using open prices, low prices, and close prices. This study comprises the use of Multi Linear Regression (MLR), Support Vector Machine (SVM) with Linear Kernel, Ploy Kernel, and RBF Kernels, Multilayer Perceptron (MLP 2, 2), and XGBOOST algorithms. To access the accuracy of results from different models, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) also been significantly calculated. Multi Linear Regression (MLR) model showed more accurate results for the prediction of day-wise high prices as compared to other algorithms. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) of MLR found 11.57 and 14.69 respectively, the lowest of all the algorithms. Traditional regression tools are not as effective as these algorithms because these algorithms can provide more accurate and precise results to serve the interests of investors. This study concluded with suggestions to the investors while deciding whether to hold the stocks or sell them on a particular day.

Keywords—Multi Liner Regression (MLR), Multi-Layer Perceptron, Stock price prediction, Support Vector Machines (SVM), XGBOOST

I. INTRODUCTION

Stock market plays prominent role in the economic development but also full with lot of volatility and risk thus it is important to identify the projected risks and movements in stock prices. Every investor want to know the behaviour of various stocks in advance so profits can be attained without more risks. Prediction of stock prices is challenging but now a day stock price prediction studies are very popular with respect to financial market. However, now a day by applying various prediction tools and techniques an investor can try to understand the reaction of stock prices. Machine learning algorithms are performing significantly in the analysing and prediction of stock prices. Uses of machine learning techniques are latest and in trend as the forecasting tools. Pabba et. al. (2023) also exhibited that the machine learning

techniques can be efficiently used to predict the stock price behaviour. Every investor takes buying and selling decision of stocks with profit intention only but on the other hand it can be stated that always profit making is not an easy task in stock market. In this concern predictive tools like as machine learning can helps a lot to the investors for forecasting the stock price. So investor can make decision regarding holding or selling the stock. Kumar et. al. (2022) favoured collaborative use of various machine learning techniques can be more useful to gain the significant results in terms of stock price prediction. This study proposes the use of machine learning techniques to predict the daily high price of the stocks by using the historical data of TCS Ltd. This high price predicted value might be significant for the investor to sell the sock on a particular day or to hold it. This study framed by dividing the data into two significant parts as training data and testing data. More and more availability of the data would be grateful to gain more meaningful results.

In this study:

- The overview of all the key technical algorithms used in the process of prediction is significantly highlighted.
- Comparative results of different algorithms are significantly studied with reference to finding of errors so more precise results can be drawn.

Current study further structured into five segments: segment two is full with the description of related studies already conducted as guidance light for the successful completion of current study. Objectives of the study and projected methodology with details of data used are exhibited in the segment three. All the analysis and results findings will be given in the segment four, stating the useful results of the current research work. While the fifth segment is disclosing the core results in terms of conclusion of whole research work.

II. LITERATURE REVIEW

Various studies exhibited about stock price prediction by using machine learning algorithms. Few of them have been evidenced in this segment of present research paper. Singh. et. al. (2023) used support vector regression (SVR), random forests and neural networks, machine learning models to predict the stock prices. Researchers found the deep learning approach are better in terms of prediction as compared to conventional models. Priya and Geetha (2022) conducted a study on use of machine learning techniques for stock price

prediction. In this study researchers had significantly used support vector and random forest after pre-processing of the data. As conclusion this study favoured machine learning models as compared to traditionally used forecasting models specifically random forest as the best forecasting model. Shinde, et. al. (2022) studied for stock price prediction by using Long Short-Term Memory (LSTM) model and regression. Open, close, low, high prices and volume traded has been used by the researcher for completion of their research work. This study concluded by increasing the size of dataset for training purpose to gain more accurate results. Pyke, et. al. (2022) studied about stock market prediction with the help of machine learning techniques. In this study researchers used regression based model and Long Short-Term Memory (LSTM) model for the stock price prediction. As conclusion researchers proposed both methods for stock value prediction. Mhasane (2022) conducted a study for stock price forecasting by using machine learning models. In this study researcher worked to find the closing price of Reliance stock by using Long Short-Term Memory (LSTM) model. This study concluded that the larger dataset can be more useful to find the gainful results with respect to stock price prediction. Kolekar, et. al. (2022) conducted a study to compare the various models of machine learning for the prediction of stock prices. This study mainly stressed on Linear regression (LR), support vector machine (SVM), Decision tree (DT) and random forecast to find the insightful results. This study resulted that newer machine learning models may give more fruitful results in forecasting. Alamir, Mesbah and Fakhr (2021) conducted study to analyse the significance of machine learning models in stock price prediction. In this study the researchers found Long Short-Term Memory (LSTM) model is more significant to find the fruitful results as compared to other techniques on the way of stock price prediction. Chen (2020) investigated on stock price prediction with the application of Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) and Support Vector Regression (SVR) machine learning techniques. Mean absolute percentage error (MAPE) has been significantly calculated to check the accuracy of various models and this study resulted that the SVR model is most suitable for prediction of the stock prices along with more accuracy of results. Nabipour et al. (2020) conducted a study to predict the stock market tendency with the application of machine learning techniques. Data has been collected from Tehran stock exchanged and decision tree, random forest, support vector classifier (SVC), Logistic regression, Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM) model applied for meaningful results. LSTM performed well as compared to others in stock prediction tendency. Kompella and Chilukuri (2019) framed a study on the use of machine learning techniques to predict the stock prices. In this study Random Forest has been significantly used and also compared with compared with logistic regression to find the meaningful results. This study resulted as the Random Forest is better than logistic regression and it can predict the stock value more accurately. Tripathy (2019) used support vector machine (SVM) model to predict closing price of S&P BSE TECK Index by using data 2008 to 2018. This study found SVM model is significantly useful in the volatile market along with high accuracy of results. Patel et. al. (2018) compared RNN and LSTM models for stock market prediction. Closing price of next day of BSE listed companies has been predicted by using open, low, close, high price and traded volume were also included in this prediction

calculation. In this study researchers found RNN had given more accurate results in price prediction as compared to LSTM model of machine learning.

III. RESEARCH METHODOLOGY

A. Objectives of the study

- To understand and compare different algorithms used in forecasting.
- To predict the daily high price of stocks with special attention on TCS ltd.
- To suggest selling or holding of stocks to the investors.

B. Description of Data

Secondary data has been considered for the completion of the present research paper. This paper is especially emphasized on TCS Ltd., with 15 years of data from 1st January, 2009 to 29th December, 2023. A total of 3714 records have been used: 12 years of data for training purposes, which includes a total of 2972 records, and 3 years of data with 742 records for testing purposes.

C. Tools for Analysis

Regression: Regression is a method used to predict discrete values. In regression, an association has to be proven among the dependent variable (say D) with independent variable (say I). This relationship is given by regression equation, $D = c + b \cdot I$, where D=estimated dependent variable value, c = constant, b =regression coefficient, and I = independent variable value. One independent and one dependent variable, is the simplest form of regression. When multiple independent variables and one dependent variable, it is called multiple regression and the regression equation, $D = c + b_1 \cdot I_1 + b_2 \cdot I_2 + b_3 \cdot I_3 + \dots + b_n \cdot I_n$.

Regression Using Support Vector Machines (SVM): Regression using SVM and the method is called Support Vector Regression (SVR). The SVR automatically translates nominal values to numeric values. The Support Vector Regression finds a paramount fit line which decreases the error of the cost function. In the training data set, support vectors or occurrences which are adjacent to the line with minimum cost. To grip the high dimensionality of the feature space, Kernel functions are used. In this paper three kernels: Linear kernel, Ploy kernel and RBF kernel significantly used.

Multi-Layer Perceptron: A MLP or multilayer perceptron is a class of feed forward artificial neural network (ANN). The Multilayer perceptrons are used to solve supervised learning problems: in the process of solving the problem, models are train on dataset of input-output pairs and teach the model about the correlation between those inputs and outputs parameters or variables. The training includes adjusting the weights or the parameters, of the implementing model to achieve low error.

XGBoost: XGBoost is a machine learning algorithm that belongs to the ensemble learning category, specifically the gradient boosting framework. It utilizes decision trees as base learners and employs regularization techniques to enhance model generalization. Known for its computational efficiency, feature importance analysis, and handling of missing values, XGBoost is widely used for tasks such as regression, classification, and for ranking.

IV. RESULTS AND DISCUSSION

To examination the models performance via experimental work we calculated correlation coefficient, mean absolute error (MAE), Root mean squared error (RMSE). The experimental work is performed by using R programming. Table-1 shows comparative result of models to predict TCS ltd. share day-wise high price.

TABLE I. TCS LTD. SHARE HIGH PRICE PREDICTION ANALYSIS BY DIFFERENT ALGORITHMS

Model	Correlation Coefficient (R)	MAE	RMSE
Multi Liner Regression (MLR)	0.9981	11.57	14.69
SVM (Linear Kernel)	0.9970	41.05	43.75
SVM (Poly Kernel)	0.9892	927	989.44
SVM (RBF Kernel)	0.9725	1021	1167.50

Model	Correlation Coefficient (R)	MAE	RMSE
MLP (2,2)	0.9788	987	1009.50
XGB	0.9905	465.60	519.72

Source: R Programming

As above shown in table 1, Multi Liner Regression (MLR) and SVM (Linear Kernel) performed better in comparison to other algorithms. Because the correlation coefficient of MLR is 0.9981 and for SVM (Linear Kernel) the correlation coefficient is 0.9970, highest as compared to other algorithms. Errors (MAE & RMSE) of Multi Liner Regression (MLR) and SVM (Linear Kernel) algorithms are also low as compared to other algorithms. R's high value and low value of MAE & RMSE proved that Multi Liner Regression (MLR) and SVM (Linear Kernel) are better algorithms than other used algorithms.

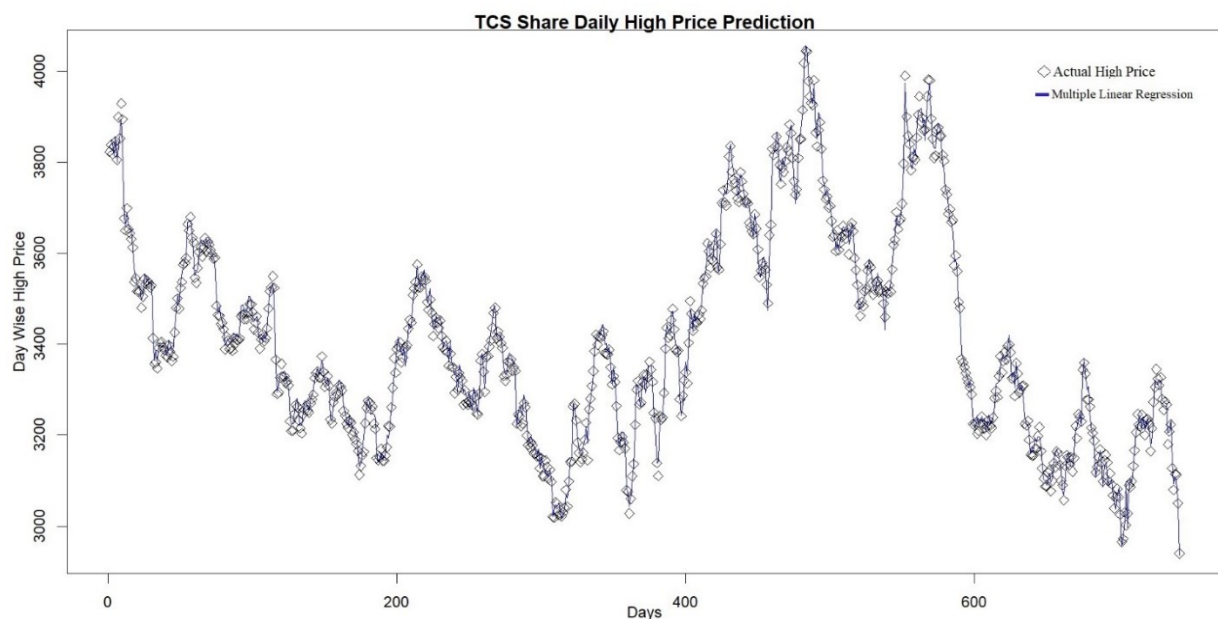


Fig. 1. Predicted price and Actual price of TCS ltd. by using MLR

Figure-1 shows the application of Multi Liner Regression (MLR) to predict the day-wise highest price. Figure 1 is

showing the predicted day to day price of share and actual price of share are very close. Hence, MLR is a useful algorithm to find the useful results for prediction purpose.

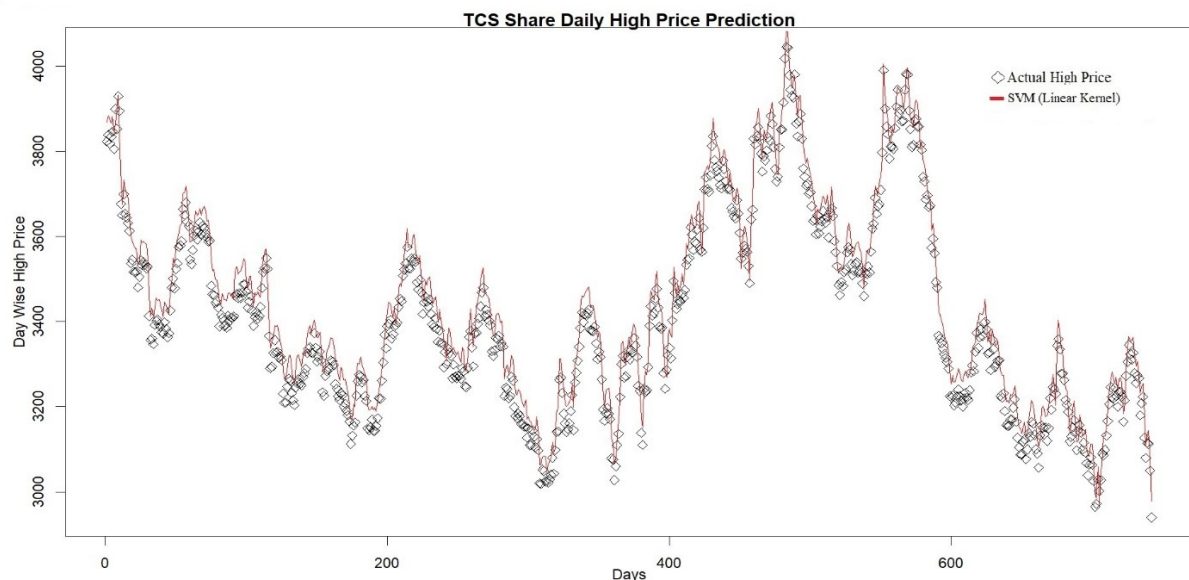


Fig. 2. Predicted price and Actual price of TCS ltd. by using SVM (Linear Kernel)

Figure-2 shows the prediction of share price by SVM (Linear Kernel) algorithm which is very close to actual price. The following figure-3 shows the comparison of predicted price by Multi Liner Regression (MLR) and SVM (Linear Kernel)

algorithm with actual price. The Multi Liner Regression (MLR) algorithm performs better than SVM (Linear Kernel) algorithm with high value of correlation coefficient and low value of errors as also evidenced by Table-1.

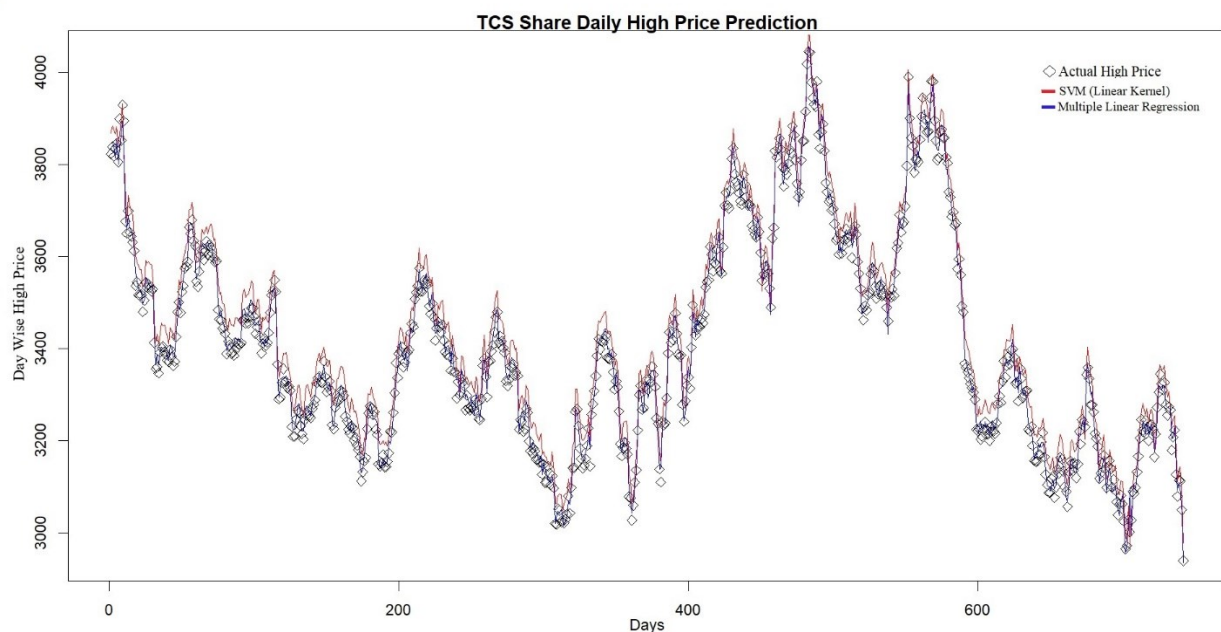


Fig. 3. Comparison of Predicted price and Actual price of TCS ltd. by using MLR and SVM (Linear Kernel)

V. CONCLUSION

Nowadays stock price prediction is a very popular research area. With the help of Machine Learning (ML) algorithms or techniques share price can be predicted by using the historical data. This study is carried out on emphasizing the TCS ltd stock prices. Day-to-day high price prediction is made in this study. Multi Liner Regression (MLR), SVM with Linear, Ploy and RBF Kernels, MLP with 2 layers and XGBOOST algorithms significantly used for the completion of this study. The MLR and SVM (Linear Kernel) performed better as compared to other algorithms. The MLR performed best among all these algorithms with low MAE=11.57 and RMSE=14.69 values, because these lowest value of errors prove more reliability on the results.

These predicted values will help to make a decision of selling of share or holding of share on a particular day.

REFERENCES

- [1] Alamir, L.A., Mesbah, S., and Fakhr, M.W. (2021). Role of machine learning in predicting stock prices: A literature survey. *Journal of Management Information and Decision Sciences*, 24(1), 1-12.
- [2] Chen, Li-Pang (2020). Using Machine Learning Algorithms on Prediction of Stock Price. *Journal of Modeling and Optimization*, 12(2), 84-99.
- [3] Kolekar, et. al. (2022). Stock price prediction comparative analysis based on machine learning algorithm. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 9(8), 532-534.

- [4] Kompella, S. and Chilukuri, K. C. (2019). Stock Market Prediction Using Machine Learning Methods. *International Journal of Computer Engineering and Technology*, 10(3), 20-30.
- [5] Kumar, et. al. (2022). Stock Market Prediction Using Machine Learning. *Journal of engineering sciences*, 13(6), 1126-1133.
- [6] Mhasane, D. (2022). Stock Market Prediction Using Machine Learning. *International Journal of Research Publication and Reviews*, 3(9), 2079-2082.
- [7] Nabipour, et. al. (2020). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access*, 8, 150199-150212.
- [8] Patel, J., Patel, M. and Darji, M. (2018). Stock Price Prediction Using RNN and LSTM. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 5(11), 1069-1079.
- [9] Pabba, et. al. (2023). Stock Market Price Prediction Using Machine Learning. *Journal of Engineering Sciences*, 14(5), 709-715.
- [10] Priya, N. V. and Geetha, S. (2022). Stock Prediction Using Machine Learning Techniques. *International Journal of Creative Research Thoughts (IJCRT)*, 10(10), d48-d60.
- [11] Pyke, et. al. (2022). Stock Price Prediction using Machine Learning with Python. *International Journal of Innovative Research In Technology*, 9(1), 833-387.
- [12] Singh, R.D., Chouhan, H. and Neema, M. (2023). Predicting Stock Prices Using Machine Learning: A Comparative Study. *International Journal of Advances in Engineering and Management (IJAEM)*, 5(4), 1812-1821.
- [13] Shindeet. et. al. (2022). Stock Prices Prediction Using Machine Learning and Deep Learning Techniques. *IOSR Journal of Engineering (IOSRJEN)*, Special issue, 15-16.
- [14] Tripathy, N. (2019). Stock Price Prediction Using Support Vector Machine Approach. *International Academic Conference on Management & Economics*, 44-59.
- [15] https://www.moneycontrol.com/stocks/histstock.php?indian_indices=4