Predict Stocks Price by Fundamental Analysis Using Machine Learning Algorithms

Mr. Shubham Dhananjay Shende
(student)
Vishwakarma University,
Pune- 411048, India
201900482@vupune.ac.in

Mr. Aniruddha Shrawankumar Singh
(student)
Vishwakarma University,
Pune- 411048, India
201900321@vupune.ac.in

Mr. Shubham Sushil Shah
(student)
Vishwakarma University,
Pune- 411048, India
201900347@vupune.ac.in

Mr. Mayuresh Mohan Shinde
(student)
Vishwakarma University,
Pune- 411048, India
202000009@vupune.ac.in

Mr. Siddhant Rahul More
(student)
Vishwakarma University,
Pune- 411048, India
201901212@vupune.ac.in

Dr. Bharati Ainapure
(Project Guide and Instructor)
Vishwakarma University,
Pune- 411048, India
bharati.ainapure@vupune.ac.in

***Abstract*— In this academic research, we make an effort to put machine learning methods for stock price prediction. Stock price forecasting uses machine learning effectively. In order to make wiser and more accurate financial decisions, machine learning methods can be examined. The scope of this research work revolves around financial ratios published by companies. This research paper put forth a stock market price forecasting technique based on the fundamental study of stocks. The technique examines the financial ratios or parameters of stocks over a specific period of time and forecasts whether they will experience a gain or loss. We found that strategies like random forest and the LSTM algorithm function best throughout the system of thinking about various strategies and factors that should be taken into account.**

Keywords*— *Stock Market, Long Short-Term Memory, Fundamental Analysis, Machine Learning, Random Forest.

I. Introduction

This The potential monetary returns are the primary reason for anticipating stock price fluctuations. Since the inception of this investing instrument, a significant amount of study has been undertaken around topic of prediction of stocks' price. Investors want to invest in companies that they believe will outperform and surpass the others or the market in an intention to gain by selling them later. Over the years, many stock prediction algorithms have been created, yet the consistency of most of these systems' actual forecast performance is still questionable. There are only a few categories in which stock prediction strategies can be classified, one of them is Fundamental analysis, which involves analyzing the underlying companies' disclosed financial accounts to make projections.

For retail individual investor it is very hard to specially allot time and carry out the analysis of stocks. There are very few applications who give predictions using fundamental analysis. So our motive is to give the users more detailed information before investing in stock. The aim of our academic project is to implement supervised machine learning algorithms to the available financial data of companies or stocks. The stock price prediction problem is interdisciplinary hence it involves applying knowledge from the domain of Machine Learning, Finance and The Stock Market.

Investors have historically used a number of well-established strategies to aid in assessing equities and forecasting price movement. Three main categories of these methods are - technical analysis, fundamental analysis, and sentiment analysis. A Trader or a technical analysts presume that stocks' price already reflects all the data available in public domain – current market price reflects everything, probability of prices following a trend is higher, and it's very likely that history repeats itself [1]. A stock's fundamental analysis is a way to determine a stock's intrinsic value by taking into account a wide range of economic aspects in relation to industry performance as well as a company's financial factors. Warren Buffet, a well-known investor around the world, is regarded as a fundamental analyst. Current technologies like Natural language processing and text analysis are used in sentiment analysis to systematically extract and detect related information/data. Sentiment analysis is used in the stock market to determine how investors feel generally about a stock or about the market as a whole. In our project, sentiment analysis will not be covered in detail because it is outside the purview of this study.

There are many variables and noises influencing price movement, hence, stock price prediction is a very complex, and challenging problem. However, for training the model and using it for predicting the prices, majority of these studies include independent variables such as historical price of stock, technical parameters, and analysis of investor's sentiment.

Universe of Stocks: This project aims to build stock prediction models. Hence, the group of stocks is known as universe of stocks [2]. In this academic project, 4 stocks – namely Axis Bank, ICICI Bank, Marico, and Infosys – are selected from the stocks listed in the Equity Segment of the National Stock Exchange of India (NSE).

Back-testing: A stock investment strategy can be tested by simulating its usage on historical stock prices, where the investment returns are calculated as if the strategy were applied at the time. Back-testing is the name of this testing technique. Back-testing is predicated on the fundamental tenet that a successful technique in the history or past is very likely to get succeed in future as well [3].

Forward-Testing: A real or fictitious investment in accordance with a plan can be used to test the strategy in the real world and gauge how well it performs when used as an investment strategy. A drawback of forward testing is the lengthy testing duration it necessitates. [3].

II. PROPOSED SYSTEM

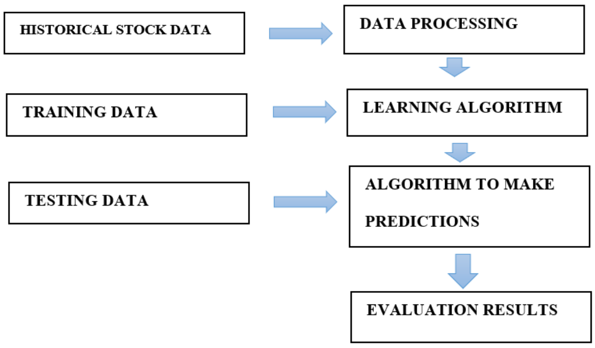

Fig. 1

A. *Data Preparation and Training*

1) *Data Collection*

A portion of the dataset for the model based on fundamental analysis was collected from "moneycontrol.com" and the remaining sector-sensitive financial ratios were collected from “investor’s presentation” available on the respective company's website.

The dataset for the model based on technical analysis was collected from “yahoofinance.com”.

2) *Filling in Missing Data Values*

Missing values will be filled by calculating the weighted average of that data feature for that particular financial year.

3) *Data Segregation*

We considered training data from FY2009 to FY2017; Validation data as from FY2017 to FY2020; And, test data from FY2020 to FY2022.

B. *Methodology*

1) *Local Learning*

There are two strategies which fall into categories as follow: local learning and global learning. Because the local learning strategy appears to perform better, we created one model for each stock for all the two machine learning algorithms or techniques.

2) *Evaluation Metrics*

Past research on the use of machine learning for stock prediction has used a variety of metrics to assess performance. Mean square error (MSE), mean absolute error (MAE), and confusion matrix are commonly used to assess the model's accuracy. We will use MSE, coefficient of determination, and accuracy (%) which are defined further in “Results and Discussion” in chapter 4.

3) *Model implementation*

Two models based on LSTM Algorithm and one model based on Random Forest Algorithm will be implemented based on “Financial Ratios”. One model based on LSTM Algorithm will be based on “Price Action” of stock. For models based on “Fundamental Analysis”, stock price at the end of each quarter will be considered. While, for model based on “Technical Analysis”, stock price at the end of each trading session will be considered.

C. *Feature Selection*

The results of various models trained with different sets of features will be analysed to determine their sensitivity to the stock price. The final model will be made up of feature sets that produce predictions that are close to the actual stock price.

D. *Machine Learning Algorithms*

1) *LSTM*

The input, output, and forget gates of the LSTM give it the innate capacity to retain information that is more pertinent and discard irrelevant information. Because of this, LSTM is a useful model for deciphering patterns across extensive timescales. The crucial aspect of LSTM is the input, which must take the form of a 3D vector (samples, time-steps, features). As a result, the input must be modified to fit this. LSTM models are among the most potent time-series models available.

2) *Random Forest*

The supervised learning algorithm Random Forest (RF) is adaptable and may be used for both classification and regression applications. During the data fitting procedure, several decision trees are constructed. When generating results for a regression problem, RF uses the arithmetic mean or aggregate value of all decision tree outputs. Voting system i.e. majority voting from the decision trees is used as the outcome for classification issues. The performance of RF can be improved by tuning a variety of hyperparameters.

Considering the reasons mentioned above these two algorithms – LSTM and Random Forest – were selected for our academic research.

III. LITERATURE SURVEY

The focus of this section is to find relevant articles for applying machine learning to stock price forecasting. First, we'll look at research that uses traditional techniques, particularly fundamental analysis, to predict stock performance. Next comes research that applies machine learning approaches using technical stock analysis. Finally, we will discuss stock price prediction techniques that combine machine learning and fundamental analysis. As a result, the results of our study are increasingly similar to those described in this section.

A. *Stock Prediction with Fundamental Analysis*

Fundamental analysis focusses on determining the intrinsic value of shares using financial indicators of a company listed on stock exchanges. The 1934 publication "Security Analysis" by Benjamin Graham and David Dodd contains the basics of fundamental analysis and is considered as the bible of fundamental analysis [4]. This book laid the theoretical foundation for value investing system. Many studies have been undertaken to formalize as well as expand the stock selection concepts presented in this book.

Piotroski developed an F-Score logistic regression model to assess the stability of a company's finances or financial condition. The “F Score” model was created using 9 financial factors that were collected from the available financial records of the company. The criteria has three categories: operational efficiency, profit, and liquidity. Using data from 1976 to 1996, Piotroski successfully back-tested the “F Score” model for stock selection [5].

In a similar way, Mohanram developed the G-Score model for stock selection. The G-Score system was created using various set of financial criteria that were collected from company’s financial data available in public domain. Profitability, naive extrapolation, and accounting conservatism make up the criteria. A significant positive correlation between the “G score” and realized returns was found when G-score model was back-tested from the FY1978 to FY2001 [6].

B. Stock Prediction with Machine Learning Based on Technical Analysis

Mostly, all the recent research that uses machine learning to predict stocks is built around technical analysis. Models based on technical analysis are widely used because they are so popular with Wall Street financial advisors and the financial press. Moreover, stock technical data is readily available in much greater volume than the data for financial or fundamental analysis. The reason is that the stocks’ prices or indices and other technical indicators are sampled every trading day or session; while, the data for fundamental or financial analysis is released once every three months. Kimoto researched how to use FNN (feedforward neural networks) for predicting the stock prices in year 1990. Various macroeconomic indicators such as the interest rate, the exchange rate, etc. as well as several technical indicators were used as inputs for their stock price prediction model.

From January 1987 to September 1989, the TOPIX index was used for buying and selling for a time period of 33 months. The results show that in terms of profitability, the neural network prediction model outperforms the buy and hold strategy. Artificial neural network (ANN), support vector machine (SVM), random forest (RF), and naive-Bayes were the 4 different Machine Learning techniques used by Patel. In this study, 10 technical indicators were used as input [7].

In this experiment, the use of daily historical data spanning 10 years for 2 stocks that belong to S & P BSE (Bombay Stock Exchange) Sensex was used. Results showed the 3 models for stock price prediction were outperformed by Rf model in terms of overall performance. Author also reckons that if the inputs are transformed from continuous value to discrete - trend deterministic - data, then performance of the prediction can be improved [8] & [9].

To increase the prediction performance, Patel advised integrating several machine learning methods. In the suggested two-stage methodology, the value of technical indicators was predicted n days in advance by using Support Vector Regression (SVR). The closing price was predicted using SVR, ANN, and RF in the second step utilizing the forecasted technical indicators. Results suggested that the two staged model performed better than the single-stage models [10].

Berkiros used data of NIKKEI and NASDAQ indices and compared the Recurrent Neural Network (RNN) and ANFIS to anticipate the trend of the following day. To avoid data swooping, data from FY1971 to FY1998 was used as train data, and from FY1998 to FY2002 was used as test data. Result suggested this ANFIS model had greater rate of return than that of the RNN model and the buy and hold strategy [11].

ANFIS model was presented by Atsalakis for forecasting the next day’s price trend. The suggested model employed the historical price as well as the price moving average as inputs. A success rate of 62.3% was observed in this ANFIS model. Results also suggest that the ANFIS model significantly outperforms the buy and hold strategy employed on each of the 5 equity stocks in terms of return on investment (ROI). Then, thr other neural-network model from past research were compared with this ANFIS model. The author asserts that this new ANFIS model has a larger chance of success than earlier models [12].

Wei proposes a neuro-fuzzy system based on KNN. The k-NN technique was used in this study to choose the “k” such that the testing input were similar to the k historical data instances. During training the model, rather than utilizing all of the data for training, KNN had been utilized to dynamically choose k instances for each prediction. Using data from 1999 to 2004 of TAIEX, the model was tested and compared to fuzzy time series and univariate neural network models. According to the results, the proposed model performed better than other baseline models in terms of RMSE [13].

C. Stock Prediction with Machine Learning Based on Fundamental Analysis

Now we move to reviewing prior research/academic studies or papers on predicting the prices of stocks using fundamental analysis by employing machine learning methods. For stock selection purpose FNN i.e. feed forward neural network model was developed by Quah and Srinivasan using quarterly fundamental financial information. The third element is average price rise over time, which is determined using weighted averages. For this experiment, quarterly stock data from Q1 FY1993 through Q4 FY1996 were used. Each stock had just 16 observations. The initial ten datapoints were used as train data while the remaining six datapoints were used as test data. A separate moving windy system approach is also made due to the limited amount of data. The moving average is calculated using three fourth as the train data and remaining as the test data. For testing the model stocks that had highest anticipated gains were chosen. Then the portfolios were evaluated according to the gains it generated.

The trial’s results show that the proposed technique may choose portfolio of stocks which may outperform and surpass the market or market indices in 10 of the 13 quarters for that testing period and produce higher gains. As authors emphasized in the concluding part of their paper, the results were primarily restricted due to the volume of data that was available. Consequently, their results are not completely conclusive [14].

By using more Data, A comparable FNN (Feed Forward Neural Network) model was developed by Lam. 364 S&P companies we used as the model's training and testing data for the period 1985 to 1995. The model's inputs consisted of sixteen financial ratios and eleven macroeconomic factors. Lam put up four research projects to look into various predictor combinations. The first three studies' inputs were financial data from year one, two and three respectively and combining all years data in subsequent years. This configuration is useful for simulating and factoring the time series effect for research. In this research project, the volume of financial data and macroeconomic data that was made use of was of three years. Lam demonstrated the outcomes of his models with various numbers of hidden layers. A portfolio was built using the first 33% of the stocks having highest anticipated gains, and its performance was evaluated. Two key observations are made as a result of the experimental results. First, experiments 1 through 3 demonstrate a steadily rising rate of return. These results show how combining Fundamental analysis with the technical analysis method of examining past patterns may boost the realized gain or percentage gain. Second, outcomes of this project shows that the model's performance is not improved by the addition of macroeconomic factors [15].

IV. RESULTS AND DISCUSSIONS

Here,
"FA LSTM Model – 1" means LSTM Model-1 implemented using Fundamental Analysis.

"FA LSTM Model - 2" means LSTM Model- 2 implemented using Fundamental Analysis.

"FA RF Model" means Random Forest Algorithm model implemented using Fundamental Analysis.

"TA LSTM Model" means LSTM model implemented using Technical Analysis.
Model Evaluation Metrics

- "MSE" :

"MSE" refers mean square error

- "CoD" :

"CoD" refers to coefficient of determination

- "Accuracy (%)" :

Accuracy (%) is calculated as follow,

Percentage= (Predicted Value / Test Value) * 100

Let's say,
TV= Total number of predicted values in range from 70% to 100%.
In predicted dataset, a value is considered in "TV" if and only if it has it's "Percentage" value in a range between 70 to 100.

$$\text{Accuracy} = \frac{\text{Total no. of values in range 70\% to 100\%}}{\text{Total no. of Values}}$$

Here, When we say "Accuracy" of a model is 70% it means that 70% of values present in the predicted dataset lie in a range from 70% to 100%.

A. *Optimizer*

Adam (Adaptive Moment Estimation) deals with first and second order momentums. The idea behind the Adam is that instead of rolling quickly only so we may reach the minimum, we should somewhat slow down to allow for a more comprehensive analysis. This optimizer gives optimal results when we are dealing a complex – interdisciplinary - problem having a lot of data and parameters. It requires less memory and is efficient.

Note: "For all the charts (of all the models) shown below, x- axis defines quarters of the testing period, and y- axis defines the prices of stocks."

B. *FA LSTM Model - 1*

LSTM Algorithm was used and implemented. Financial ratios sensitive respective stocks were used for analysis.

"Axis Bank" stock model was the worst performer while, "Infosys" stock model gave the best performance according to our model evaluation metrics.

"Adam" optimizer gave more accurate results as compared to "SGD" and "Relu" optimizers.

In the charts below, orange line (---) shows predicted value and blue line shows (---) actual values.

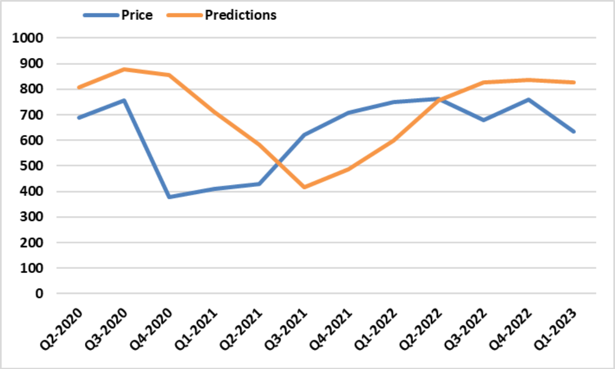

Fig. 2 Axis Bank

Here we can observe that the prediction line and the line of actual price are parallel from Q2-2020 to Q3-2020. Further, a decline in the graph lines can be observed. The contradiction of lines can be clearly seen between graph lines from Q4-2020 to Q42021 -the line of prediction and the actual price have directions of slopes opposite. As both the lines have deviated and varying slopes for most of the time, the selected stock "Axis Bank" gave the worst performance for LSTM Model 1.

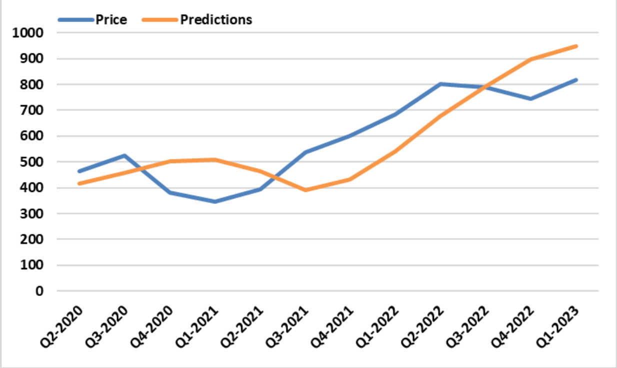

Fig. 3 ICICI Bank

Here lines of prediction and lines of actual value have nearly similar slopes for most of the time period.

Contradiction can be observed at Q3-2020 to Q2-2021, wherein, the lines of prediction and actual price show a completely opposite slope, which results in a huge difference between the predicted and actual price and thereby affecting the accuracy of prediction. Similar contradiction can again be observed at Q2-2022 to Q4-2022.

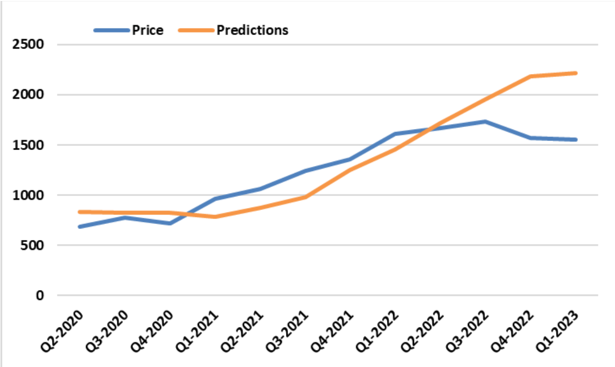

Fig. 4 Infosys

Here the slope of both, the prediction line and the line of actual price have similar slopes for most of the time. Contradiction can still be observed from Q4-2020 to Q1-2021, and again from Q3-2022 to Q1-2023. Even though the graph line had varying slopes, "Infosys" stock gave the best performance for this LSTM Model.

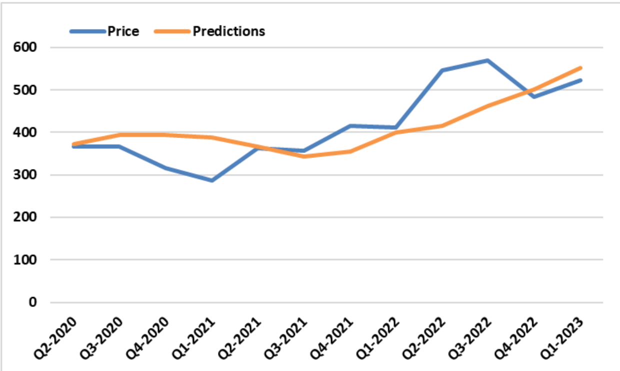

Fig. 5 Marico

Here, contradiction can be observed from Q2-2020 TO Q2-2021.The graph of stock "Marico" showed many instances of varying slopes. As a result, the performance of this model can be termed as average.

C. FA LSTM Model – 2

Once again LSTM Algorithm was implemented in a different way. Financial ratios sensitive respective stocks were used for analysis.

"Marico" stock model was the best performer while, "Axis Bank" stock model gave the worst performance according to our model evaluation metrics.

For this model also, "Adam" optimizer gave more accurate results as compared to "SGD" and "Relu" optimizers.

In the charts below, red line (---) shows predicted value and green line shows (---) actual values.

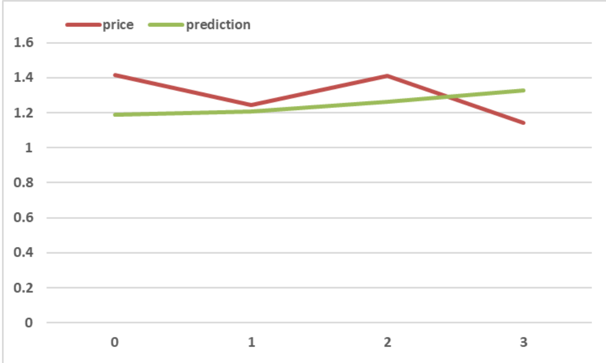

Fig. 6 Axis Bank

Here a decline in the slope of line of actual price can be seen from 0.0 to 1.0. Simultaneously, an incline in the line of prediction can be observed at the same points leading to contradiction. There are 2 such observations the slopes of lines vary from each other very much leading to a very poor performance in prediction.

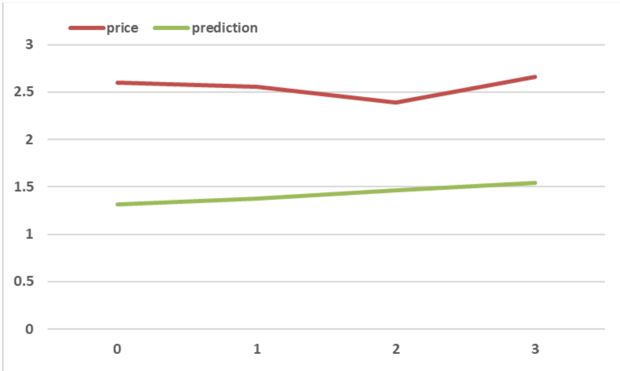

Fig. 7 ICICI Bank

The line of actual price showed a decline in slope from 0.0 to 2.0 and an increase from 2.0 to 3.0. Whereas, the line of prediction showed a continuous incline. Hence, we can say that the model performed pretty average for the stock "ICICI Bank" as compared to other stocks.

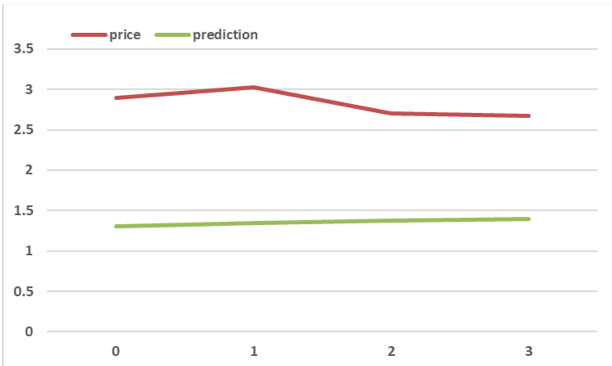

Fig. 8 Infosys

Both the lines have similar slopes from 0.0 to 1.0 . From 1.0 contradiction can be observed. At 2.0 slight decline in the slope of line of actual price can be observed, whereas at the same point, a slope close to 0 can be observed in the line of prediction. Here too, the performance of the model for this stock can be termed as average.

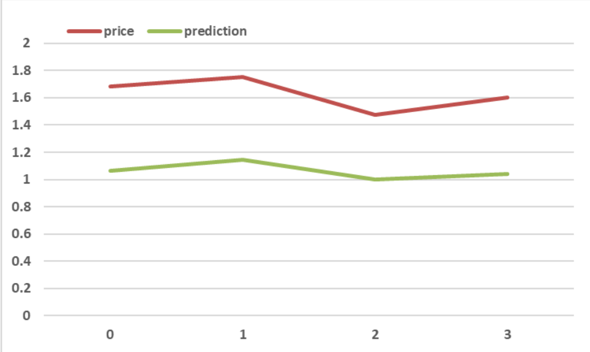

Fig. 9 Marico

The Marico stock had similar slopes for both, the line of prediction and the line of actual price at every instance. Hence, we can say that the stock performed well for this LSTM Model.

D. FA RF Model

Random Forest Algorithm was used to model the stock prices. Financial ratios sensitive respective stocks were used for analysis.

"Marico" stock model was the worst performer while, "Infosys" stock model gave the best performance according to our model evaluation metrics.

In the charts below, orange line (---) shows predicted value and blue line shows (---) actual values.

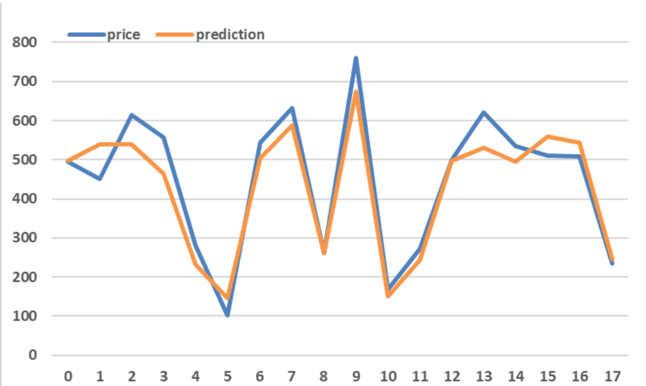

Fig. 10 Axis Bank

The line of prediction and actual price can be observed with similar slopes most of the time. There are instances where contradiction can be observed. Hence, we can say that the stock "Axis Bank" performed average for this model.

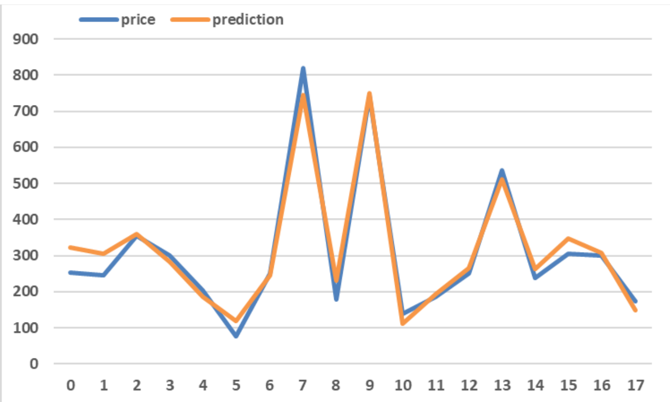

Fig. 11 ICICI Bank

No contradiction can be observed between the lines of prediction and the line of the actual price, but the slopes vary slightly. Hence, the performance of this stock can be said to be above average for this model.

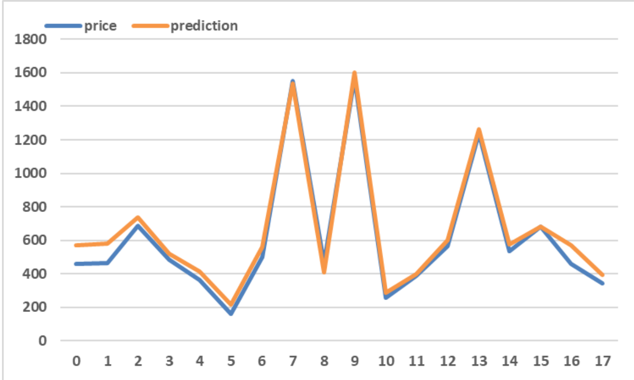

Fig. 12 Infosys

"Infosys" stock performed well compared to other stock models for this RF Model. No contradiction can be observed between the line of prediction and the line of the actual price. There are instances where the slope of lines is exactly the same. Hence, the stock "Infosys" performed well compared to the other stocks for this model.

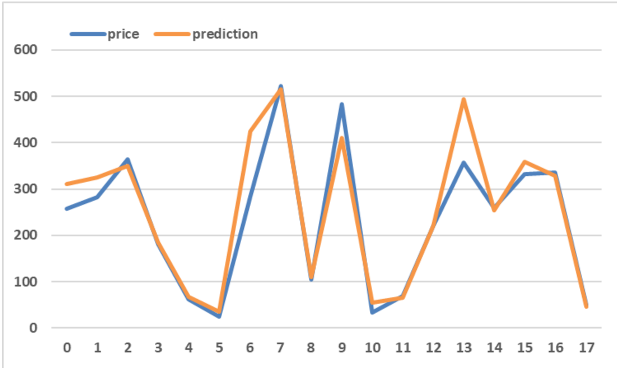

Fig. 13 Marico

"Marico" stock performed poorly as compared to other stock models for this RF Model. No contradiction can be observed between the lines of prediction and the line of the actual price, but a huge difference in the slopes of both lines can be observed making its performance very poor for this model.

E. TA LSTM Model – 1

LSTM Algorithm was implemented to model stock prices; But in this model, the price action of the stock was considered. Every trading session stock's "close price" was considered. Hence, in other words, we can say technical analysis was used in this Machine Learning Model.

"ICICI Bank" stock model was the best performer while, the "Marico" stock model gave the worst performance according to our model evaluation metrics.

In the charts below, orange line (---) shows predicted value and blue line shows (---) actual values.

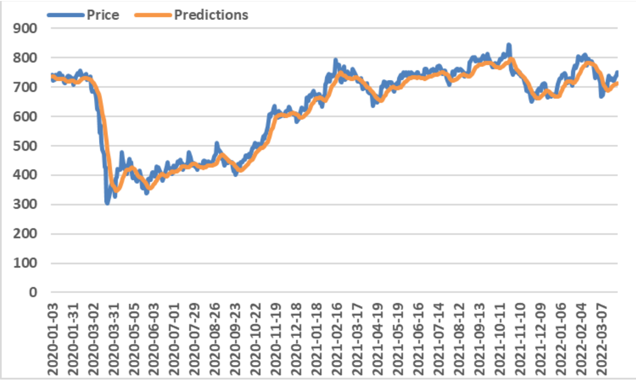

Fig. 14 Axis Bank

Although, the predicted line and the actual line follow each other closely, there are many instances where over a short-range due to deviation in the predicted value and the actual value can lead to loss to investors for short time frame. Hence, “Axis Bank” stock gave an average performance as compared to other stock models for this LSTM Model.

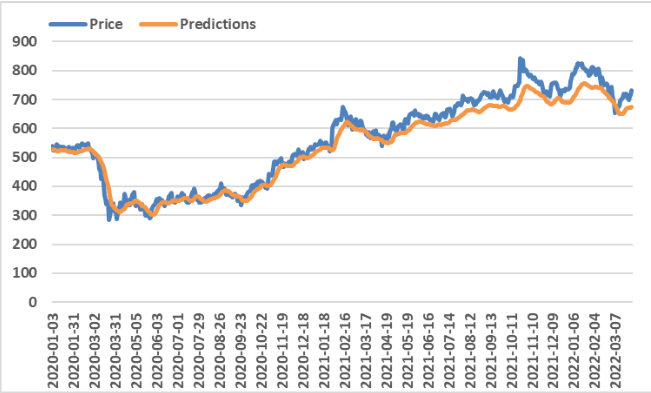

Fig. 15 ICICI Bank

As we can see visually that the predicted line and the actual line follow each other very closely. Hence, “ICICI Bank” stock performed very well as compared to other stock models for this LSTM Model.

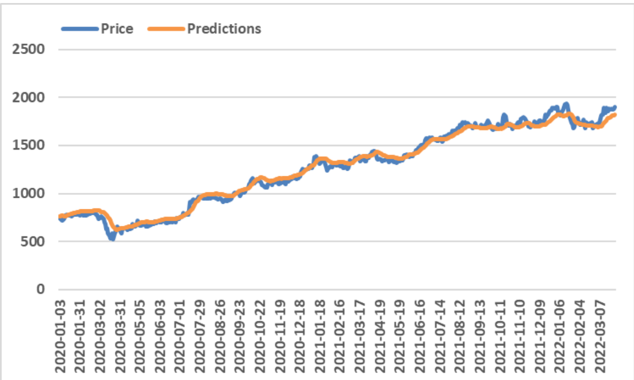

Fig. 16 Infosys

As seen in the “Axis Bank” model, a similar trend was observed here also. The predicted line and the actual line follow each other closely, there are many instances where over a short-range due to deviation in the predicted value and the actual value can lead to loss to investors for short time frame. So, “Infosys” stock gave average performance as compared to other stock model for this LSTM Model.

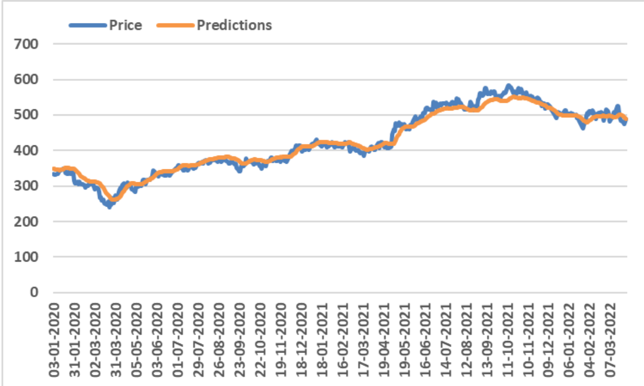

Fig. 17 Marico

“Marico” stock gave a poor performance as compared to other stock models for this LSTM Model.

F. *Overall Analysis*

Stock price prediction based on technical analysis implemented using LSTM Algorithm gave the best performance as compared to other models. While stock price prediction based on fundamental analysis implemented using LSTM (FA LSTM Model - 2) was the worst performer.

The closer MSE is to 0, the better the model predicts. CoD should be 1 for a perfect prediction and a Negative value of CoD represents the opposite view as shown by the actual value.

Below Tables shows Model Evaluation Metrics of all (four) stocks for all (four) models.

Table 1

FA LSTM 1	MSE	CoD	Accuracy
Marico	4487.57	-0.65	75
Axis Bank	46936.38	-1.10	16.67
ICICI Bank	20659.66	0.57	50
Infosys	110844	0.66	41.67

Table 2

FA LSTM 2	MSE	CoD
Marico	0.31	-118
Axis Bank	0.03	-8.53
ICICI Bank	1.07	-191.13
Infosys	1.69	-781.44

Table 3

FA RF	MSE	CoD	Accuracy
Marico	2763.5	0.89	44.45
Axis Bank	2901	0.89	61.12
ICICI Bank	1343.64	0.96	38.89
Infosys	3459.17	0.98	12

Table 4

TA LSTM 1	MSE	CoD	Accuracy
Marico	396.16	0.93	49.73
Axis Bank	1488.08	0.91	69.30
ICICI Bank	545.07	0.98	58.17
Infosys	4088.61	0.97	89.95

1) FA LSTM-1

Marico: Here, mean square error is observed to be 4487.57. The coefficient of determination is -0.65. The model resulted in a 75% accuracy for the prediction of the stock price on test data.

Axis Bank: Here, mean square error is observed to be 4687.57. The Coefficient of determination is -1.10. Negative value of CoD represents the opposite view as shown by the actual value. The model resulted in a 16.67% accuracy for prediction of future stock price, which can be termed as very poor.

ICICI Bank: Here, mean square error is observed to be 20659.66. The Coefficient of determination is 0.57. The model resulted in a 50% accuracy for prediction of future stock price, which can be said to be an average performance of the model.

Infosys: Here, mean square error is observed to be 110844. Coefficient of determination is 0.66. The model model resulted in a 41.67% accuracy for prediction of future stock price.

2) FA LSTM-2

For all the FA LSTM-2 Models,

MSE values are calculated w.r.t. scaled values of prices of stocks from 0 to 1; And, all the values of CoD are way beyond the expected value i.e. not closer to 1. So, the accuracy(%) of the respective stock's model is not computed.

Marico: Here, the Mean Squared Error is observed to be 0.31. Coefficient of determination is -118.

Axis Bank: Here, the Mean Squared Error is observed to be 0.03. The Coefficient of determination is -8.53.

ICICI Bank: Here, the Mean Squared Error is observed to be 1.07. coefficient of determination is -191.13.

Infosys: Here, the Mean Squared Error is observed to be 1.69. The Coefficient of determination is -781.44.

3) FA RF

Marico: Here, mean square error is computed to be 2763.5. Coefficient of determination is 0.89.The model resulted in a 44.45% .

Axis Bank: Here, mean square error is computed to be 2901. Coefficient of determination is 0.89. model resulted in a 61.12% accuracy.

ICICI Bank: Here, mean square error is computed to be 1343. Coefficient of determination is 0.66. model resulted in a 38.89% accuracy.

Infosys: Here, mean square error is computed to be 3459.17. Coefficient of determination is 0.66. model resulted in a 12% accuracy.

4) TA LSTM:-

Marico: Here, mean square error is observed to be 396.16. Coefficient of determination is 0.93 model resulted in a 49.73% accuracy.

Axis: Here, mean square error is observed to be 1488.08. Coefficient of determination is 0.91. model resulted in a 69.30% accuracy.

ICICI Bank: Here, mean square error is observed to be 545.07. Coefficient of determination is 0.98. model resulted in a 58.17% accuracy.

Infosys: Here, mean square error is observed to be 4088.61. Coefficient of determination is 0.97. model resulted in a 89.95% accuracy.

V. Conclusion, Observations and Future Work

Machine Learning models of predicting the stock prices using the fundamental/financial ratios analysis (FA) gave poor performance as compared to the model for stock price prediction using the technical analysis (TA).

A. Accuracy in FA models can be increased as follow:

- Company declare their quarterly results within 45 days, after the quarter has ended. Hence, in dataset stock price must be selected of the day when the company announces their results.
- Generally, euphoria of the result of a company can be seen in stock price, i.e. price of the stock upto 3 days before the company announces the result and upto 2 days after the results are announced. Hence, one may consider taking the average of stock price of these 5 days to include reaction of market w.r.t. the results published by company.
- Another way could be to calculate weighted average of prices that stock has been in the respective quarter.
- For our academic project, we choose dataset from FY2009 to FY2022, to increase the accuracy dataset

for more longer period can be taken into account, say, FY1988 to FY2022.

TA model's dataset had stock price of every trading session happened in FY2009 to FY2022; Hence, model had many datapoints i.e. 2808 datapoints to learn, validate and test. This resulted in more accuracy of TA model.

B. *Future Work*

Let's assume a possibility that an "optimal" model exists; However, given the short amount of time and resources allocated to this project, we could not assert that we have produced a flawless "optimal" model in this academic research.

Moreover, there are numerous ways to dig deep into this subject. First of all, because corporations only release their financial data on a quarterly basis, the small amount of data available in this project significantly limits the models' ability to predict outcomes. Addition of more data might enhance both the effectiveness of our model and the accuracy of our findings. More algorithms, including various neural network versions, could be tested.

Additionally, several approaches of feature selection can be experimented and analyzed. After employing feature selection techniques, choosing how many features to retain is a somewhat individual choice. To further enhance the efficacy of our feature selection, we may experiment with varying the quantity of the most crucial qualities. We could try including sentiment analysis in our model as well.

ACKNOWLEDGMENT

We would like to express our gratitude and regards to Dr. Bharati Ainapure Ma'am for guiding us throughout the life cycle of this academic project work. We are grateful to all the authors whose work we have cited referred. We offer our sincere thanks to all who directly or indirectly helped us in building this academic research paper.

REFERENCES

[1] Christina Majaski, "Technical Analysis: Fundamental Vs. Technical Analysis", Investopedia, [Online]. Available: https://www.investopedia.com/university/technical/techanalysis2.asp
[2] J. Chen, "Investopedia-Universe of Securities," [Online]. Available: https://www.investopedia.com/terms/u/universeofsecurities.asp
[3] Jean Folger,"Backtesting and Forward Testing: The Importance of Correlation"[Online].
Available: https://www.investopedia.com/articles/trading/10/backtesting-walkforward-important-correlation.asp
[4] B. Graham and D. Dood, Security Analysis, McGraw-Hill Book Co., 1934.
[5] J. D. Piotroski, "Value investing: the use of historical financial statement information to separate winners from losers," Journal of Accounting Research, vol. 38, pp. 1-41, 2000.
[6] P. S. Mohanram, "Separating winners from losers among low book-to-market stocks using financial statement analysis," Review of Accounting Studies, vol. 10, no. 2-3, pp. 133-170, 2005.
[7] T. Kimoto, K. Asakawa, M. Yoda and M. Takeoka, "Stock market prediction system with modular neural networks," in IEEE, IJCNN International Joint Conference on Neural Networks, 1990.
[8] J. Patel, S. Shah, P. Thakkar and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," Expert Systems with Applications, vol. 42, no. 1, pp. 259-268, 2015.
[9] J. Patel, S. Shah, P. Thakkar and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," Expert Systems with Applications, vol. 42, no. 4, pp. 2162-2172, 2015.
[10] E. Chong, C. Han and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," Expert Systems with Applications, vol. 83, pp. 187-205, 2017.
[11] S. D. Bekiros and D. A. Georgoutsos, "Evaluating direction-of-change forecasting: Neurofuzzy models vs. neural networks," Mathematical and Computer Modelling, vol. 46, no. 1-2, pp. 38-46, 2007.
[12] G. S. Atsalakis and K. P. Valavanis, "Forecasting stock market short-term trends using a neuro-fuzzy based methodology," Expert Systems with Applications, vol. 36, no. 7, pp. 10696-10707, 2009.
[13] C.-C. Wei, T.-T. Chen and S.-J. Lee, "A k-NN Based Neuro-Fuzzy System for Time Series Prediction," in 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2013.
[14] T.-S. Quah and B. Srinivasan, "Improving returns on stock investment through neural network selection," Expert Systems with Applications, vol. 17, pp. 295-301, 1999.
[15] M. Lam, "Neural network techniques for financial performance prediction: integrating fundamental and technical analysis," Decision Support Systems, vol. 37, no. 4, pp. 567-581, 2004.