Auto Sector Stock Price Trend Prediction using Decision Tree

Manish M. Goswami

Abstract: The auto sector stock price trend is based on many national and international uncertain factors. It is challenging to predict the impact of such a factor on the stock price trend as the impact of the same factor varies at different points of time. In this research work, we are predicting the auto sector stock price trend using patterns in the historical data using a machine learning method.

Keywords: Decision Tree, Supervised Learning, Unsupervised Learning

I. INTRODUCTION

T he stock market prediction has generated a lot of interest in academia and business world. Aim of this is to address the question: How much historical data of stock's price is useful to meaningfully predict the future price of stock? [1, 2, 3, 4]. Efficient Market Hypothesis (EMH) and the random walk theory were useful in early analysis of stock exchange prediction. These early models were of the opinion that stock prices are driven by latest information instead of present/past price and hence it is not possible to predict them. Thus, stock market relies on a random walk and their prediction accuracy is less than 50% [4]. However many are proposing opposite to EMH and random walk hypotheses. To some extent prediction of stock market is possible as per these studies [5, 6] and so, raising a question mark on the EMH as underlying assumptions. Warren Buffet as seen by business community is able to beat the S&P index [7, 8] in a consistent manner serving a practical indicator about the market prediction. In the next section, various methods that are widely used in the prediction of the stock market are discussed.

II. PREDICTION METHODS

Prediction methods are of three categories described below in detail:-

A. Fundamental Analysis

Fundamental Analysts takes into consideration of corporate that highlights the stock for that corporate. Here

company's past performance is judged along with the credibility of its accounts. Fundamental analyst helps to create several performance ratios with determining whether stock is valid or not, like the P/E ratio. Famous example of fundamental analyst is Warren Bufett. Fundamental analysis operates on the concept of human tendency of making capital to progress and if an organization progresses, it is gifted with additional capital making a surge in a stock price. Fundamental analysis is popular among fund managers because it provides affordability, objectivity and uses publically officered information like financial statement analysis. In addition to bottom-up company analysis, it also follows top down analysis in this order: first global economy, then country analysis followed by sector analysis. At the end company level analysis completes the process.

B. Technical analysis

In technical analysts the future price of a stock is supposed to rely on the (potential) tendency of the past price (time series analysis form) and is independent of company's fundamentals. Besides the patterns like head and shoulders or cup and saucer, a technique such as the exponential moving average (EMA) is used in this analysis. Japanese rice merchants who believed to have been first developers of Candle stick patterns are now commonly used by technical analysts.

C. Machine Learning

Machine Learning empowers computers the ability to imbibe human learning [9]. Machine learning algorithms broadly are of two types: Supervised Learning and Unsupervised Learning. In training a machine learning model, both an algorithm and the data is used to make model learn its parameters from the given training data. Learning can be supervised or unsupervised described below:-
Supervised Learning - In supervised learning, a set of input data and their associated output data is used to train the machine learning model. In other words, it is equivalent to approximate function $y = f(x)$, where x is input and y is its output. This is supervised learning as it requires training dataset with correct labels so that it acquires learning capability. Based on the output variable regression and classification can be termed are subgroups of supervised learning. Regression task outputs a variable as a continuous one while classification outputs a categorical variable such as color, shape type, etc. In practice, supervised learning algorithms are used in maximum machine learning applications.

Retrieval Number: F9882038620/2020©BEIESP
DOI:10.35940/ijrte.F9882.059120
Journal Website: www.ijrte.org

1131

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Unsupervised Learning - Machine learning algorithm if trained using input data without output data then it is a unsupervised algorithms. Here an attempt is made to know the underlying patterns of the input data thereby discovering patterns. Unsupervised learning algorithms can be divided further based on two approaches. They are Clustering and Association. In clustering, an algorithm is interested in knowing inherent clusters or groups present in the data. K-means algorithm is based on this approach. If algorithm tries to predict the future purchase then it is an example of association rule mining. Example is apriority algorithm.

Various Artificial intelligence approaches are used for prediction which delivers results with improved accuracy and hence are widely used in predicting future stock prices. Neural networks are one of them widely used in finance domain. In Neural network past and future value of time series data are related with each other nonlinearly. The success of the ANN is directly proportional to determining whether input data set is nonlinearly related or not without requiring the details of relation between the input and the output. This is an clearly an improved technique over traditional technical analysis techniques where basic concepts of trends, historical prices and volume, price patterns and oscillators are used by investors for taking decisions regarding investments to asses future stock prices. Successful Deployment of Neural networks is seen in intelligent trading systems [5, 6, 10].

III. METHODLOGY

This section describes, in brief, the Decision Tree Method, which is used in this work to solve the classification problem. Decision Tree is a tree data structure used for classification. It is a Supervised Machine Learning. Decision Tree consists of
1. Nodes: Test for one or attribute values.
2. Edges/ Branch: associated with possible value of decision and join to the next node or leaf.
3. Leaf nodes: associated with outcome (or class labels)
4. Non leaf nodes: associated with decision.

In this algorithm, classification is divided into a set of choices about each component in the feature vector. A usefulness of this

algorithm is that it helps to visualize and interpret the output of order of visiting nodes to come at the required classification. Internally, decision trees examine our input data and look. After examining input data samples the tree is then constructed for the best possible nodes/values to split on using algorithms like CART or ID3 making predictions.

IV. PROPOPSED SYSTEM

In this research, we proposed an automated trading system that will generate a trading signal and will take trading action accordingly i.e. it will execute the trading order based on the signal generated. The main logic of our system is that it is based on the previous day's n−1 trading activities; it will predict the next day n stock price trend. We have trained Decision tree classifiers using historical data. The input features are Open, Close, High, Low prices and Volume; the output can be UP, DOWN or NO trend. The experimentation is performed on the stock from the National Stock Exchange, India viz Tata Motors and State Bank of India. The

experimental data is shown in Table 1. This study includes the transaction cost of 0.10% on both sides i.e. sell and buys sides.

V. EXPERIMENTATION

This section describes, in brief about the experimental data and experiments performed.
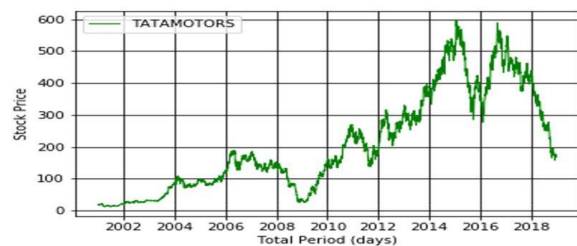
A. Experimental Data

The experimental data is divided into training and testing data as shown in Table-I. Fig.1, Fig. 2 are the daily closing price series of the experimental stocks, Tata motors and SBI respectively.

Table-I: Experimental Stock

Stock Name	Time	Total Period	Training Period	Testing Period
Tata Motors	Start	2001-01-01	2001-01-01	2015-05-25
	End	2018-12-31	2015-05-22	2018-12-31
SBI	Start	2001-01-01	2001-01-01	2015-05-25
	End	2018-12-31	2015-05-22	2018-12-31

Table- II: Performance of the proposed model on four experimental stock from the index stock NIFTY

Stock Name	Precision	Recall	F1 score	Accuracy score	Sum of daily % Return (proposed model)	Sum of daily % return(Buy and Hold Model)
SBI	0.45	0.45	0.44	0.45	44.7	22.6
Tata Motors	0.48	0.47	0.46	0.48	19.6	-88.3

Fig.1. Daily close price of experimental stock Tata Motors

**Fig.2. Daily close price of experimental stock SBI**

B. Experimentation

We have experimented with our proposed model on the stock from the Indian stock market. We created an automated agent which will take a trading decision based on the previous trading day's trading activities.

It will take its trading position at the opening of the market and will close the trading position at a closed price. Trading position can be long, short or no position.

VI. RESULTS AND DISCUSSION

The proposed model recommends the day trader the stocks to take a long position, short position, or take no position among the set of stocks provided. The performance of the proposed model for this classification problem is recorded in TABLE 2 in terms of Precision, Recall, F1-score, Accuracy on four stocks mentioned in Table-I.

We backtested the proposed model, by taking trading decisions recommended by the proposed model for day trading. If the signal provided by the proposed model is taking a long position, we bought the stock at an open price and sold it at a close price. Similarly, if the signal provided by the proposed model is taking a short position, we sold the stock at an open price and bought it at a close price. The performance of this trading strategy is in terms of the sum of the daily % Return and we compared the proposed model with the Buy-and-Hold trading strategy. The comparison of the performance of the proposed model and the Buy-and-Hold stock trading strategy on SBI stock is shown in Fig.3. Similarly, Fig.4 depicts a similar comparison on Tata Motors.
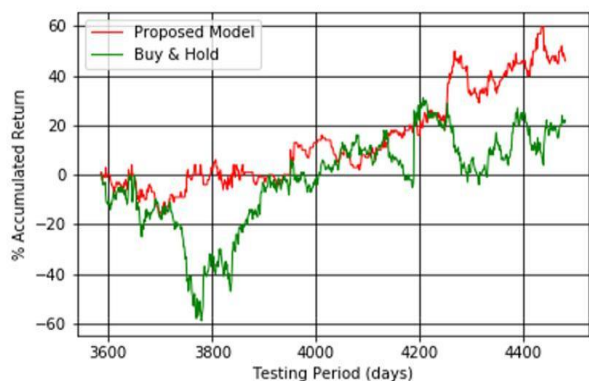
Fig. 3. Performance of the proposed model and Buy-and-Hold trading strategies on SBI stock in terms of the sum of daily % Return.
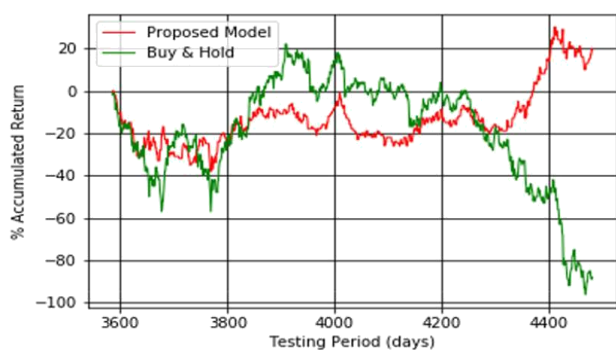
Fig.4. Performance of the proposed model and Buy-and-Hold trading strategies on TATAMOTORS stock in terms of the sum of daily % Return.

VII. CONCLUSION

In the stock market, the stock price quickly reflects any event or news related to the stock in its price. Our work is based on this principle, so we can extract meaningful information from the recent trading activity of the stock. We have predicted the next day's stock price trend from previous day trading activity. We recorded previous day trading activity for all trading days whenever the stock price trend varies significantly up or down from historical data. Extracted trading rules for long and short positions from recorded data in terms of ten features. We selected features which are useful for short term stock price trend. These trading rules are generated using the Random Forest method. The experimental result in TABLE 2 indicates that the prediction accuracy of our proposed model is comfortably acceptable. We created and backtested the day stock trading strategy using the recommendation of the proposed model and its result outperforms the Buy-and-Hold stock trading strategy. So the proposed model is better than the Buy-and-Hold trading strategy. In future more or different features can be used. We have used Random Forest to generate trading rules, more machine learning methods can be tested for future improvements.

REFERENCES

1. Y. Shynkevich, T. M. McGinnity, S. Coleman, and A. Belatreche, "Stock price prediction based on stock-specific and sub-industry-specific news articles," in *2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2015.
2. T. T. Vu, S. Chang, Q. T. Ha, and N. Collier, "An experiment in integrating sentiment features for tech stock prediction in twitter," in *Proceedings of the workshop on information extraction and entity analytics on social media data*, pp. 23–38, 2012.
3. Y. Zhai, A. Hsu, and S. K. Halgamuge, "Combining news and technical indicators in daily stock price trends prediction," in *International symposium on neural networks*, pp. 1087–1096, Springer, 2007.
4. J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of computational science*, vol. 2, no. 1, pp. 1–8, 2011.
5. T. Kimoto, K. Asakawa, M. Yoda, and M. Takeoka, "Stock market prediction system with modular neural networks," in *1990 IJCNN international joint conference on neural networks*, pp. 1–6, IEEE, 1990.
6. L. Wang and Q. Wang, "Stock market prediction using artificial neural networks based on hlp," in *2011 Third International Conference on Intel li-gent Human-Machine Systems and Cybernetics*, vol. 1, pp. 116–119, IEEE, 2011.
7. M.-Y. Chen, D.-R. Chen, M.-H. Fan, and T.-Y. Huang, "International transmission of stock market movements: an adaptive neuro-fuzzy inference system for analysis of taiex forecasting," *Neural Computing and Applications*, vol. 23, no. 1, pp. 369–378, 2013.
8. R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The azfin text system," *ACM Transactions on*
9. *Information Systems (TOIS)*, vol. 27, no. 2, pp. 1–19, 2009.
10. L. R. Marwala, *Forecasting the stock market index using artificial intelligence techniques*. PhD thesis, 2010.
11. S.H. Poon and C. W. Granger, "Forecasting volatility in financial markets: A review," *Journal of economic literature*, vol. 41, no. 2, pp. 478–539, 2003.
12. H. Yang, L. Chan, and I. King, "Support vector machine regression for volatile stock market prediction," in *International Conference on Intel ligent Data Engineering and Automated Learning*, pp. 391–396, Springer, 2002.
13. K.j. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1-2, pp. 307–319, 2003.

AUTHORS PROFILE

Dr. Manish M. Goswami, PhD is an associate professor in the department of Information Technology, Rajiv Gandhi College of Engineering and Research, Nagpur. He has 20+ years of teaching experience and has administrative experience of 10 years in the academic field.