International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 03 | March - 2022 ISSN: 2582-3930

A Hybrid Machine Learning Technique for Stork Market Prices Prediction

Miss Akifa Arshi, M Tech, Dept. Of CSE, Sreyas institute of engineering and technology, Nagole, Hyderabad, T.S, India, 500 068.

Mrs Joshi Padma N, Associate Professor, Dept. Of CSE , Sreyas institute of engineering and technology, Nagole, Hyderabad, T.S, India, 500 068.

Abstract:

Stock markets in financial domain play crucial role in economy of nations in the contemporary world. There are many stakeholders who depend on the prediction of stock prices for making buying and selling decisions. In this context, it became important to deal with automatic stock prices prediction. Since manual observation is not possible, machine learning (ML) techniques are widely used for prediction of stock prices movement. Many ML algorithms came into existence. However, their performance largely depends on the quality of training data as they are based on unsupervised learning. In this paper, a hybrid ML framework is proposed to have both feature selections to improve quality of training and the leveraging forecasting performance. The proposed framework supports different techniques such as Linear Regression, XG Boost Regression and Gradient Boost. A prototype application is built using Python data science platform. Experimental results revealed that the three prediction models are providing high level of accuracy. Linear Regression showed 99.989% accuracy, GBoost regression 99.981 and XGBoost Reggressor 0.99969. From the results, it is ascertained that the proposed framework is useful for efficient forecasting of stock prices movements.

Keywords – Stock Market Prediction, Machine Learning, Linear Regression, GBoost Regression, XGBoost Reggressor

1. INTRODUCTION

Stock markets in financial domain play crucial role in economy of nations in the contemporary world. There are many stakeholders who depend on the prediction of stock prices for making buying and selling decisions. In this context, it became important to deal with automatic stock prices prediction. Since manual observation is not possible, machine learning (ML) techniques are

widely used for prediction of stock prices movement. Many ML algorithms came into existence. However, their performance largely depends on the quality of training data as they are based on unsupervised learning. There are many contributions from existing research. Yoo *et al*. [1] explored different ML techniques used for prediction of stock markets. The found that the prediction is not trivial and challenging. They used different approaches in ML and studied their performance. They found that there is need for efficient ML techniques in order to have better prediction of stock prices. They also considered different global events and factors that are causes of stock market prices movements. They envisaged that it is still required to improve the prediction models for better performance.

Patel *et al*. [2] worked in BSE datasets that are associated with many stock tickers in India. They used different techniques like SVM, Random Forest and Neural Networks. They found that many of the techniques in ML are not capable of understanding time-series data in financial domain. They also suggested to make use of pre-processing in order to have better quality data for training. Feature selection also is advocated besides the improvement of algorithms for improving prediction performance. They applied different metrics to ascertain performance of ML models they investigated. From [3]-[15], there are different ML models used for prediction of stock market prices movement. From the literature, it is understood that there are many ML techniques that contributed to forecasting of stock prices. However, the problem is to have an ideal combination of feature selection and ML in order to have better performance. Provided a stock market dataset, building a hybrid framework that consists of feature selection followed by ML algorithm to predict stock prices is the problem considered. Our contributions in this paper are as follows.

1. A hybrid ML framework is proposed to have both feature selections to improve quality of training and the leveraging forecasting performance.
2. The proposed framework supports different techniques such as Linear Regression, XG Boost Regression and Gradient Boost.
3. A prototype application is built using Python data science platform.

The remainder of the paper is structured as follows. Section 2 reviews literature on different prediction models for stock markets. Section 3 presents the proposed methodology and underlying algorithm. Section 4 provides experimental results while Section 5 concludes the paper and gives future scope.

2. RELATED WORK

This section reviews literature on different existing methods. Yoo *et al*. [1] explored different ML techniques used for prediction of stock markets. The found that the prediction is not trivial and

challenging. They used different approaches in ML and studied their performance. They found that there is need for efficient ML techniques in order to have better prediction of stock prices. They also considered different global events and factors that are causes of stock market prices movements. They envisaged that it is still required to improve the prediction models for better performance. Patel *et al*. [2] worked in BSE datasets that are associated with many stock tickers in India. They used different techniques like SVM, Random Forest and Neural Networks. They found that many of the techniques in ML are not capable of understanding time-series data in financial domain. They also suggested to make use of pre-processing in order to have better quality data for training. Feature selection also is advocated besides the improvement of algorithms for improving prediction performance. They applied different metrics to ascertain performance of ML models they investigated.

Roberts and Morrissey [3] investigated moving targets in stock markets and provided their insights with ML techniques. Since stock market prices are subjected to different influences and their prices movement cannot be easily predicted, they proposed a methodology to understand different factors and influencing events in order to have better approach to deal with stock market data. Their experiments showed that there is further need for understanding stock market in connection with sentiments. Deepak *et al*. [4] also studied ML techniques where supervised ML is used. They also opined that mere supervised learning with training data is not sufficient for accurate prediction. There is need for incorporating different ideas, factors and also sentiment analysis. They leverage performance of prediction models with some feature selection approaches. It is understood that feature selection has something to do with quality of training thereby influences the performance of prediction models. The real time datasets are used and performance in prediction of different models are evaluated. ANN based models are found to be more useful in prediction. In other words, such models are able to ascertain data distributions correctly.

Hegazy *et al*. [5] used different optimizations for stock market prices prediction. The investigated on different ML model and optimized them. They found that both buyers and sellers need to have required intelligence in the market place to succeed. Towards this end they implemented models based on PSO and SVM and optimized them further. Different pre-processing approaches and NLP techniques are employed to leverage potential of the algorithms used to predict stock prices. The prediction of next day stock price and the prediction based on given time interval such as 10 days etc. are investigated with optimized ML models. Dash and Kishore [6] investigated on ANN based prediction models that make use of certain approach that mimics human brain. Neurons are used to model the data distributions and trading

decisions are extracted. The prediction of decisions is found to be complex as they depend on so many factors. However, buy and hold decisions are based on the market predictions whether they are done intuitively or with scientific approach considering ML models. The proposed network based model provides a decision support system that is used to have more accurate predictions that lead to expert decision making. With different ML models, they provided a profitable stock trading decision approach that is the novel aspect in their research. From [3]-[15], there are different ML models used for prediction of stock market prices movement. From the literature, it is understood that there are many ML techniques that contributed to forecasting of stock prices. However, the problem is to have an ideal combination of feature selection and ML in order to have better performance. Provided a stock market dataset, building a hybrid framework that consists of feature selection followed by ML algorithm to predict stock prices is the problem considered.

3. PROPOSED SYSTEM Stock markets in financial domain play crucial role in economy of nations in the contemporary world. There are many stakeholders who depend on the prediction of stock prices for making buying and selling decisions. In this context, it became important to deal with automatic stock prices prediction. Since manual observation is not possible, machine learning (ML) techniques are widely used for prediction of stock

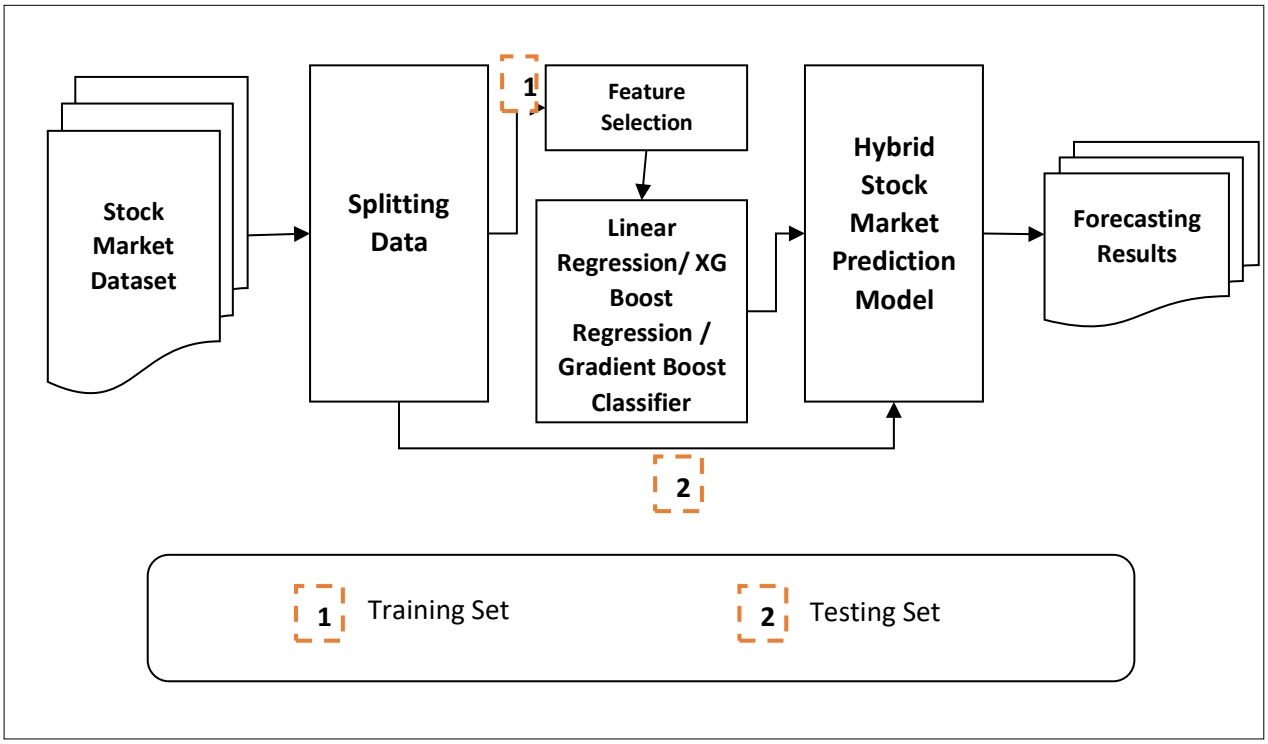

Figure 1: Proposed ML based framework for stock market prices prediction

movement. Many ML algorithms came into existence. However, their performance largely depends on the quality of training data as they are based on unsupervised learning. In this paper, a hybrid ML framework is proposed to have both feature selections to improve quality of training and the leveraging forecasting performance. The proposed framework supports different techniques such as Linear Regression, XG Boost Regression and Gradient Boost. A prototype application is built using Python data science platform.

As presented in Figure 1, a hybrid ML framework is proposed and implemented for forecasting stock prices. The proposed framework supports different techniques such as Linear Regression, XG Boost Regression and Gradient Boost. It has provision for feature selection and ML techniques for improving performance. Linear Regression, XG Boost Regression and Gradient Boost are ML models suitable for forecasting stock prices. They are applied appropriately in the proposed framework. The framework is realized with an algorithm named Hybrid Machine Learning for Stock Prices Prediction (HML-SPP) which takes stock market datas*et al*ong with ML pipeline (Linear Regression, Gradient Boosting Regression, XGBoost) as input. The feature selection plays crucial role in the proposed framework as it improves quality in training leading to improved prediction performance.

Algorithm: Hybrid Machine Learning for Stock Prices Prediction (HML-SPP)
Inputs: stock market dataset *D*
machine learning models pipeline M
(Linear Regression, Gradient Boosting Regression, XGBoost)
Output: Prediction results *P*

1. Start
2. Initialize models map *M*
3. Initialize results vector *R*
4. Initialize ensemble map E
5. (*T1*, *T2*)←PreProcess(*D*)
6. *F*←FeatureSelection(*T1*)
7. For each model m in *M*
8. Train the model *m* using *F*
9. Fit the model t for T2
10. Add results to R
11. End For
12. Display R
13. End

Algorithm 1: Hybrid Machine Learning for Stock Prices Prediction (HML-SPP)

As presented in Algorithm 1, it takes stock market dataset and ML pipeline as inputs and performs ML

based forecasting of stock prices movement. It has mechanisms to split data into 80% training and 20% testing besides training the classifiers and testing the data for forecasting based on the trained knowledge. It also has feature selection process before training in order to improve quality of dataset for training.

4. EXPERIMENTAL RESULTS

A prototype is made using Python to demonstrate proof of the concept. An excerpt from dataset used for empirical study is as follows.

Date	Symbol	Series	Prev	Open	High	Low	Last	Close	VWAP	Volume	Turnove
03-01-	RELIAN	EQ	233.05	237.5	251.7	237.5	251.7	251.7	249.37	4456424	1.11E+1
04-01-	RELIAN	EQ	251.7	258.4	271.85	251.3	271.85	271.85	263.52	9487878	2.5E+14
05-01-	RELIAN	EQ	271.85	256.65	287.9	256.65	286.75	282.5	274.79	2683368	7.37E+1
06-01-	RELIAN	EQ	282.5	289	300.7	289	293.5	294.35	295.45	1568228	4.63E+1
07-01-	RELIAN	EQ	294.35	295	317.9	293	314.5	314.55	308.91	1987097	6.14E+1
10-01-	RELIAN	EQ	314.55	317.4	318.7	305.3	306.65	308.5	312.35	1341705	4.19E+1
11-01-	RELIAN	EQ	308.5	307.95	310.95	283.85	288.5	288.5	296.4	1254432	3.72E+1
12-01-	RELIAN	EQ	288.5	289	305	282.15	304.7	301.7	294.57	1210950	3.57E+1
13-01-	RELIAN	EQ	301.7	306	316.4	304.1	309.75	311.85	311.79	1707604	5.32E+1
14-01-	RELIAN	EQ	311.85	309.5	321.65	309.5	317	316.3	316.17	1346059	4.26E+1

Table 1: Shows an excerpt from dataset

As presented in Table 1, it has an excerpt from the stock market dataset used for empirical study. It has attributes that provide historical data that is used to have ML based learning for forecasting.

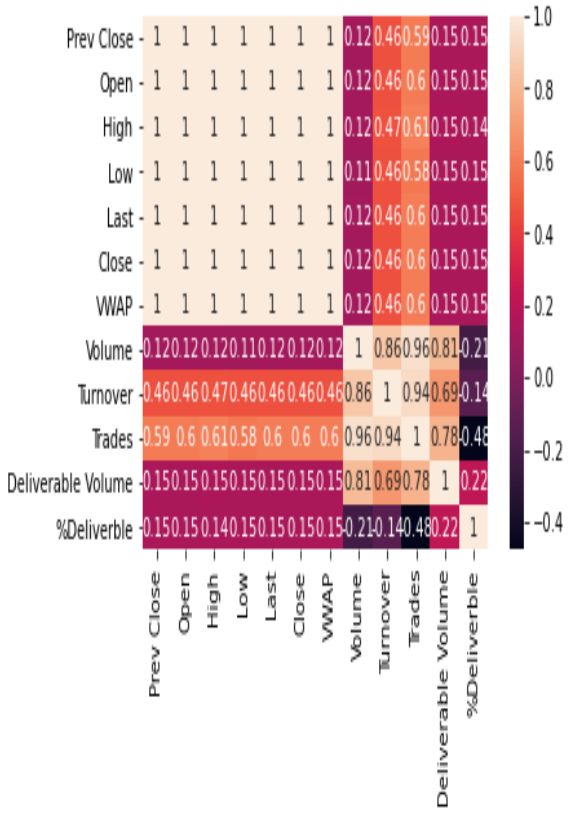

Figure 2: Shows heatmap for data visualization

As presented in Figure 2, it shows the heatmap for data visualization. It helps in understanding dynamics of stock market data.

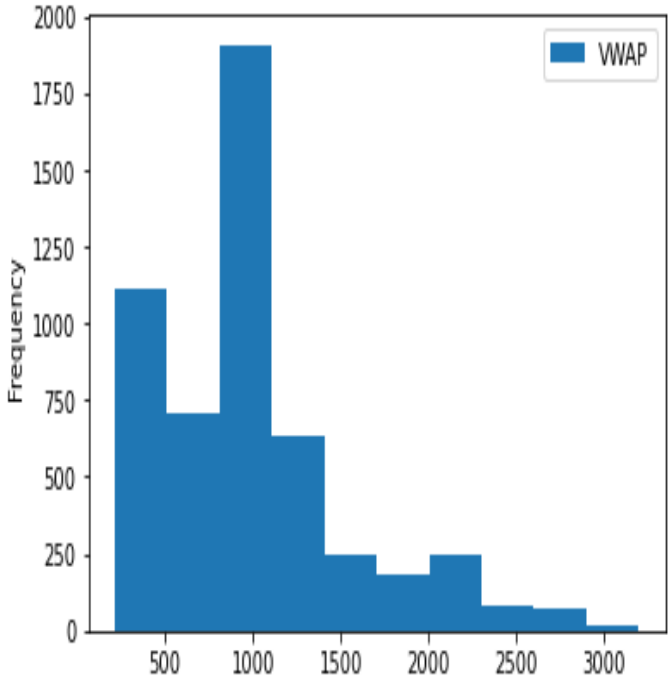

Figure 3: Shows predictions pertaining to Volume Weighted Average Price (VWAP)

As presented in Figure 3, the predictions are provided in terms of VWAP which is useful tool for traders to understand the trends and make well informed decisions in connection with in or out of the trade.

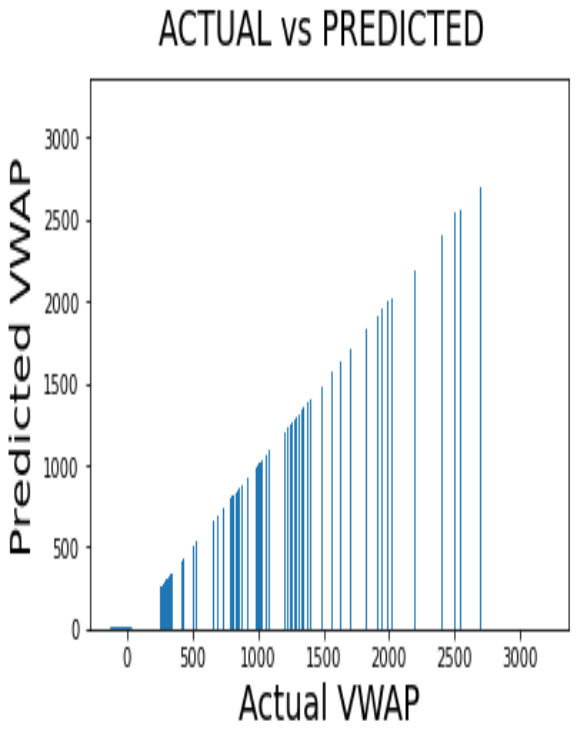

Figure 4: The VWAP predictions compared with actual values

As presented in Figure 4, the VWAP predicted values and actual values are provided in order to ascertain the performance of the ML based hybrid prediction model.

Stock Prices Prediction Model	**Accuracy (%)**
Linear Reggression	0.99989
GBoost regression	0.99981
XGBoost Reggressor	0.99969

Table 2: Performance comparison

As presented in Table 2, the accuracy of the three prediction models is provided to understand their performance.

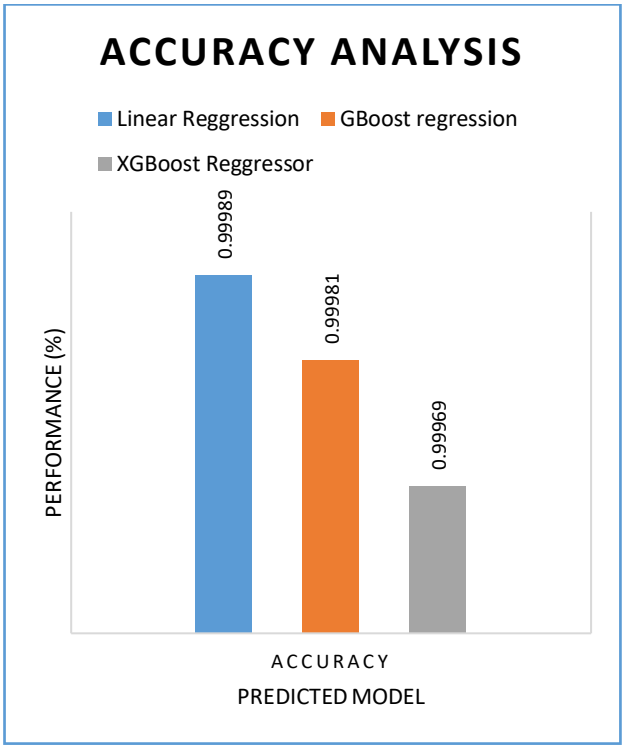

Figure 5: Shows performance comparison in terms of accuracy

As presented in Figure 5, the prediction models and their accuracy performance are provided. Each model has shown different level of accuracy. Experimental results revealed that the three prediction models are providing high level of accuracy. Linear Regression showed 99.989% accuracy, GBoost regression 99.981 and XGBoost Reggressor 0.99969. From the results, it is ascertained that the proposed framework is useful for efficient forecasting of stock prices movements.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a methodology for hybrid ML framework for stock prices movement prediction. Feature selection and supervised ML techniques are used for efficient forecasting of stock prices. The implementation includes different techniques such as Linear Regression, XG Boost Regression and Gradient Boost. A prototype application is built using Python data science platform. Experimental results revealed that the three prediction models are providing high level of accuracy. Linear Regression showed 99.989% accuracy, GBoost regression 99.981 and XGBoost Reggressor 0.99969. From the results, it is ascertained that the proposed framework is useful for efficient forecasting of stock prices movements. In future, it is interesting to explore deep learning methods along with ML techniques for further improvement in prediction performance.

References

[1] Yoo, P.D. and Kim, M.H.; Jan, T. (2005). International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06) - Machine Learning Techniques and Use of Event Information for Stock Market Prediction: A Survey and Evaluation. , 2, p835–841.

[2] Patel, Jigar; Shah, Sahil; Thakkar, Priyank and Kotecha, K (2015). Predicting stock market index using fusion of machine learning techniques. Expert Systems with Applications, 42(4), p2162–2172.

[3] Banik, Shipra; Khodadad Khan, A. F. M. and Anwer, Mohammad (2014). Hybrid Machine Learning Technique for Forecasting Dhaka Stock Market Timing Decisions. Computational Intelligence and Neuroscience, 2014,p1–6.

[4] Raut Sushrut Deepak, Shinde Isha Uday and Dr. D. Malathi. (2017). MACHINE LEARNING APPROACH IN STOCK MARKET PREDICTION. International Journal of Pure and Applied Mathematics. 115 (8), p71-77.

[5] Osman Hegazy, Omar S. Soliman and Mustafa Abdul Salam. (2013). A Machine Learning Model for Stock Market Prediction. International Journal of Computer Science and Telecommunications . 4 (12), p17-23.

[6] Dash, Rajashree and Dash, Pradipta Kishore (2016). A Hybrid Stock Trading Framework Integrating Technical Analysis with Machine Learning Techniques. The Journal of Finance and Data Science, p1-25.

[7] Waqar, Muhammad; Dawood, Hassan; Guo, Ping; Shahnawaz, Muhammad Bilal and Ghazanfar, Mustansar Ali (2017). [IEEE 2017 13th International Conference on Computational Intelligence and Security (CIS) - Hong Kong, Hong Kong (2017.12.15-2017.12.18)] 2017 13th International Conference on Computational Intelligence and Security (CIS) - Prediction of Stock Market by Principal Component Analysis. , p599–602.

[8] Henrique, Bruno Miranda; Sobreiro, Vinicius Amorim; Kimura, Herbert (2019). Literature Review: Machine Learning Techniques Applied to Financial Market Prediction. Expert Systems with Applications, p1-69.

[9] Gurav, U., & Sidnal, N. (2018). Predict Stock Market Behavior: Role of Machine Learning Algorithms. Advances in Intelligent Systems and Computing, p383–394.

[10] Chou, Jui-Sheng; Nguyen, Thi-Kha (2018). Forward Forecast of Stock Price Using Sliding-window Metaheuristic-optimized Machine Learning Regression. IEEE Transactions on Industrial Informatics, p1–11.

[11] Patel, Jigar; Shah, Sahil; Thakkar, Priyank and Kotecha, K (2015). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications, 42(1), p259–268.

[12] Vijh, Mehar; Chandola, Deeksha; Tikkiwal, Vinay Anand and Kumar, Arun (2020). Stock Closing Price Prediction using Machine Learning Techniques. Procedia Computer Science, 167 p599–606.

[13] Kumar, D., Sarangi, P. K., and Verma, R. (2021). A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings. P1-5.

[14] Sukhman Singh, Tarun Kumar Madan, Jitendra Kumar and Ashutosh Kumar Singh. (2019). Stock Market Forecasting using Machine Learning: Today and Tomorrow. IEEE, p738-745.

[15] Banik, Shipra; Khan, A. F. M. Khodadadand Anwer, Mohammad (2012). [IEEE 2012 15th International Conference on Computer and Information Technology (ICCIT) - Chittagong, Bangladesh (2012.12.22-2012.12.24)] 2012 15th International Conference on Computer and Information Technology (ICCIT) - Dhaka stock market timing decisions by hybrid machine learning technique. , p384–389.