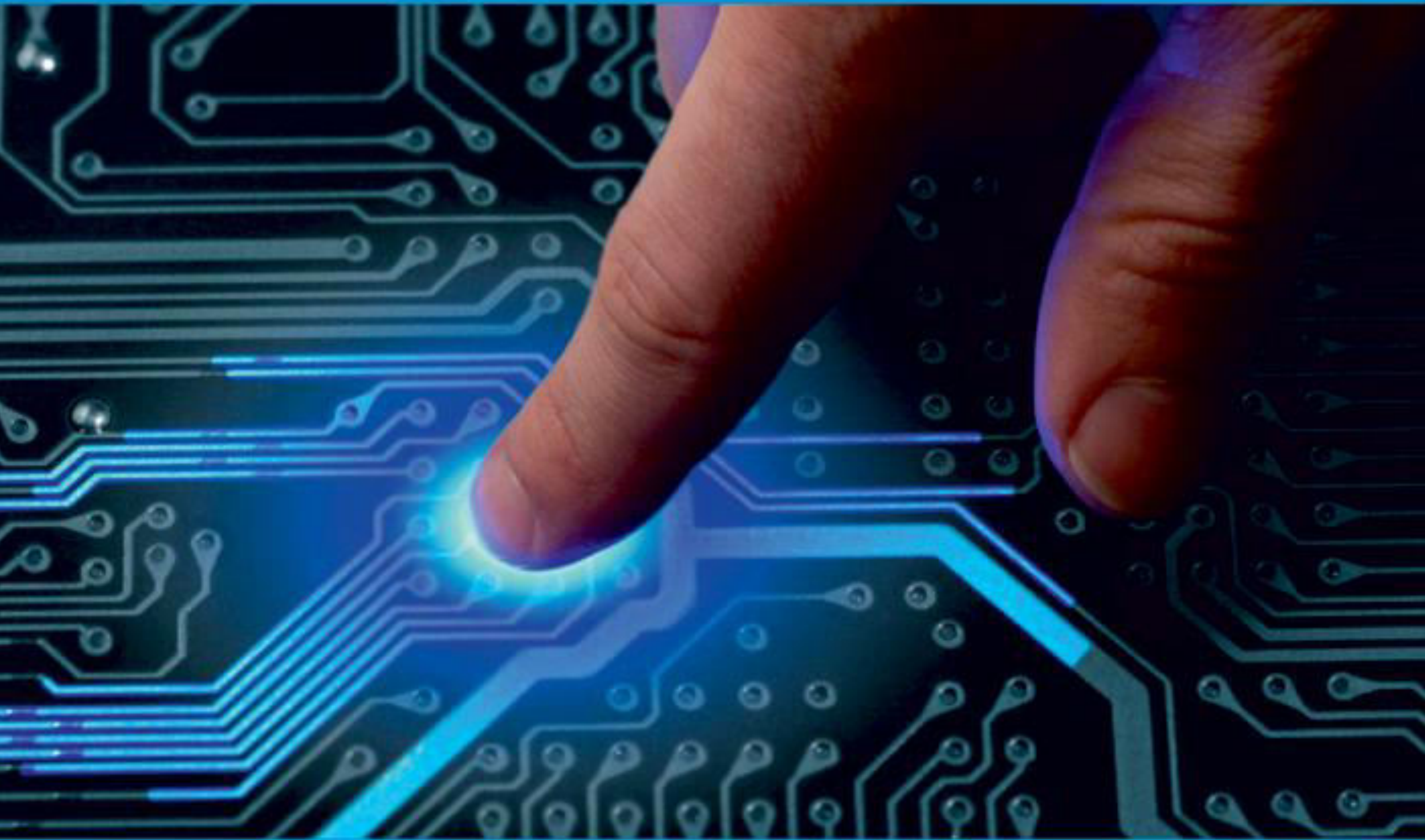
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006058 |

A Novel Artificial Intelligence Enabled Model for Stock Prices Prediction

SUBHANI SHAIK[1], Dr.JAIDEEP GERA[2]

PG Scholar, Department of CSE, ST. Marys Group of Institutions Guntur, AP, India. [1]

Associate Professor, Department of CSE, ST. Marys Group of Institutions Guntur, AP, India[2]

ABSTRACT: Stock prices prediction using machine learning models is an important research area. As the ML models exploit training data that is historical in nature, they are capable of improving prediction performance. In the literature, many ML models are found to be useful and contributing to this kind of research. However, it is understood that there is need for combination of feature selection and ML for better performance. Towards this end, in this paper, a hybrid ML framework is proposed to have both feature selectionand machine learning to improve quality of training and leverage 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 –Artificial Intelligence, Machine Learning, Stock Market Prediction, GBoost, XGBoost, Linear Regression

I. 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. This paper is aimed at building a hybrid ML framework for improving stock market predictions.

Kara *et al*. [3] used support vector machines (SVM) and artificial neural network (ANN) for knowing stock prices index movement over a period of time. Tsaih *et al.* [4] proposed a hybrid AI system to forecast stock index with the use of S&P 500 stock dataset. Polamuri *et al.* [5] proposed a GAN based hybrid prediction model for stock market prices prediction. This model exploits Stock-GAN with data augmentation to improve performance. Nti *et al*. [13] made technical analysis of different stock prediction models while SVM is used by Karathanasopoulos *et al*. [14] for stock prices prediction. Choudhry *et al*. [15] defined a hybrid prediction model for prediction purposes. Qui *et al*. [16] used ANN towards learning from historical data of stock markets and make predictions about future prospects. Ticknor *et al.* [17] proposed a Bayesian regulated ANN for forecasting stocks. Yudong *et al*. [18] used BCO and BP models combined for forecasting stock markets.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.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006058 |

1. We proposed a hybrid ML framework with feature selection and supervised learning for prediction of stock market prices.
2. We implemented different ML models along with a feature selection for improving performance.
3. A prototype application is built in order to evaluate the framework with the underlying models.

The remainder of the paper is structured as follows. Section 2 reviews literature on the existing methods used for stock market prediction. Section 3 presents the proposed ML based framework. Section 4 presents results of empirical study. Section 5 concludes the paper and gives directions for future work.

II. RELATED WORK

This section reviews literature on the stock prices prediction models. Gholamiangonabadi *et al.* [1] used hybrid methods to examine technical indicators in stock market prices prediction. Kim *et al.* [2] proposed a hybrid approach with neural networks and genetic algorithms to discover temporal patterns from stock markets. Kara *et al.* [3] used support vector machines (SVM) and artificial neural network (ANN) for knowing stock prices index movement over a period of time. Tsaih *et al.* [4] proposed a hybrid AI system to forecast stock index with the use of S&P 500 stock dataset. Polamuri *et al.* [5] proposed a GAN based hybrid prediction model for stock market prices prediction. This model exploits Stock-GAN with data augmentation to improve performance. Mallikarjuna and Prabhakara [6] explored different forecasting methods useful for stock market predictions. Chung *et al.* [7] used Long Short Term Memory (LSTM) optimized and genetic algorithm combination for prediction of stock markets. Alavi and Gandomi [8] used ANN coupled with a hybrid method beside simulated annealing for prediction of principal ground motion parameters pertaining to stock markets. Yun *et al.* [9] proposed hybrid GA-XGBoost algorithm that is supported by three-stage feature engineering for improving stock prices prediction. Asadi *et al.* [10] made a hybrid of data pre-processing and Levenberg–Marquardt neural networks that are evolutionary in nature for stock market prediction. Bahrammirzaee [11] investigated on different AI models and ANNs for intelligent financial applications. Shah *et al.* [12] explored different prediction models existing for stock forecasting.

Nti*et al.* [13] made technical analysis of different stock prediction models while SVM is used by Karathanasopoulos *et al.* [14] for stock prices prediction. Choudhry *et al.* [15] defined a hybrid prediction model for prediction purposes. Qui *et al.* [16] used ANN towards learning from historical data of stock markets and make predictions about future prospects. Ticknor *et al.* [17] proposed a Bayesian regulated ANN for forecasting stocks. Yudong *et al.* [18] used BCO and BP models combined for forecasting stock markets. Chen *et al.* [19] combined KNN model and feature weighted SVM for stock index movement predictions. Liu *et al.* [20] used random time strength function and Legendre neural network combination for forecasting stock prices. 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.

III. PROPOSED SYSTEM

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. The ML models used in this framework are ensemble models. It does mean that, they do have mechanisms to deal with multiple tree in the process of prediction. Thus they are capable of improving prediction performance. Moreover, in this paper, they are trained with dataset that is subjected to feature selection. Therefore, it ensures that the quality of training data is increased. This will reflect in the improved prediction performance.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006058 |

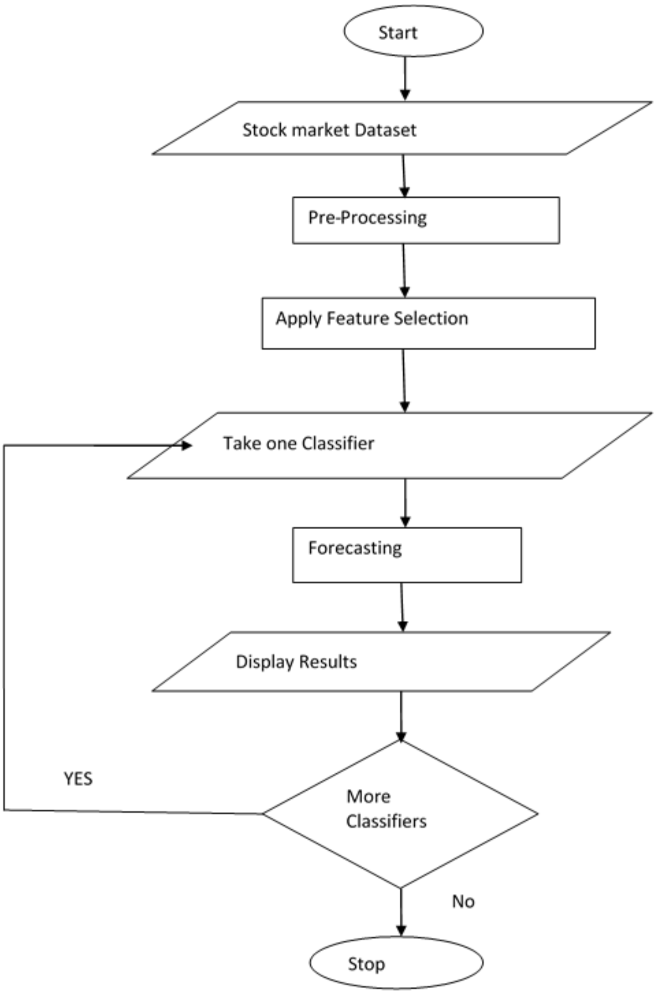


Figure 1: The functional flow of the proposed system

As presented in Figure 1, the stock market data is taken as input and it is subjected to pre-processing. Then feature selection is used to identify usefulfeatures for forecasting stock prices. This will have positive impact on the quality of training. Then it has an iterative process that deals with the forecasting process with each prediction model. Three models are used and all of them are ensemble models that have provision to ensemble multiple internal models to improve performance in forecasting.

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

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006058 |

5. (*T1*, *T2*)←Pre Process(*D*)
6. *F*←Feature Selection(*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. Compute loss function
13. Display R
14. End

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

As presented in Algorithm 1, a hybrid ML model is proposed and implemented. The Algorithm takes stock market dataset as input along with ML models pipeline. The algorithms create an ensemble map and the data is divided into 80% training and 20% testing. Afterwards features are selected from T1 that is training set. Then each model is trained with features and the classification is made to detect brain stroke samples.

IV. 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 Close	Open	High	Low	Last	Close	VWAP	Volume	Turnover
03-01-2000	RELIANCE	EQ	233.05	237.5	251.7	237.5	251.7	251.7	249.37	4456424	1.11E+14
04-01-2000	RELIANCE	EQ	251.7	258.4	271.85	251.3	271.85	271.85	263.52	9487878	2.5E+14
05-01-2000	RELIANCE	EQ	271.85	256.65	287.9	256.65	286.75	282.5	274.79	26833684	7.37E+14
06-01-2000	RELIANCE	EQ	282.5	289	300.7	289	293.5	294.35	295.45	15682286	4.63E+14
07-01-2000	RELIANCE	EQ	294.35	295	317.9	293	314.5	314.55	308.91	19870977	6.14E+14
10-01-2000	RELIANCE	EQ	314.55	317.4	318.7	305.3	306.65	308.5	312.35	13417057	4.19E+14
11-01-2000	RELIANCE	EQ	308.5	307.95	310.95	283.85	288.5	288.5	296.4	12544322	3.72E+14
12-01-2000	RELIANCE	EQ	288.5	289	305	282.15	304.7	301.7	294.57	12109507	3.57E+14
13-01-2000	RELIANCE	EQ	301.7	306	316.4	304.1	309.75	311.85	311.79	17076042	5.32E+14
14-01-2000	RELIANCE	EQ	311.85	309.5	321.65	309.5	317	316.3	316.17	13460592	4.26E+14

Table 1: Shows an excerpt from the stock market dataset used for empirical study

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006058 |

As presented in Table 1, the stock market dataset contains the required data from which training is made to forecast the stock values.

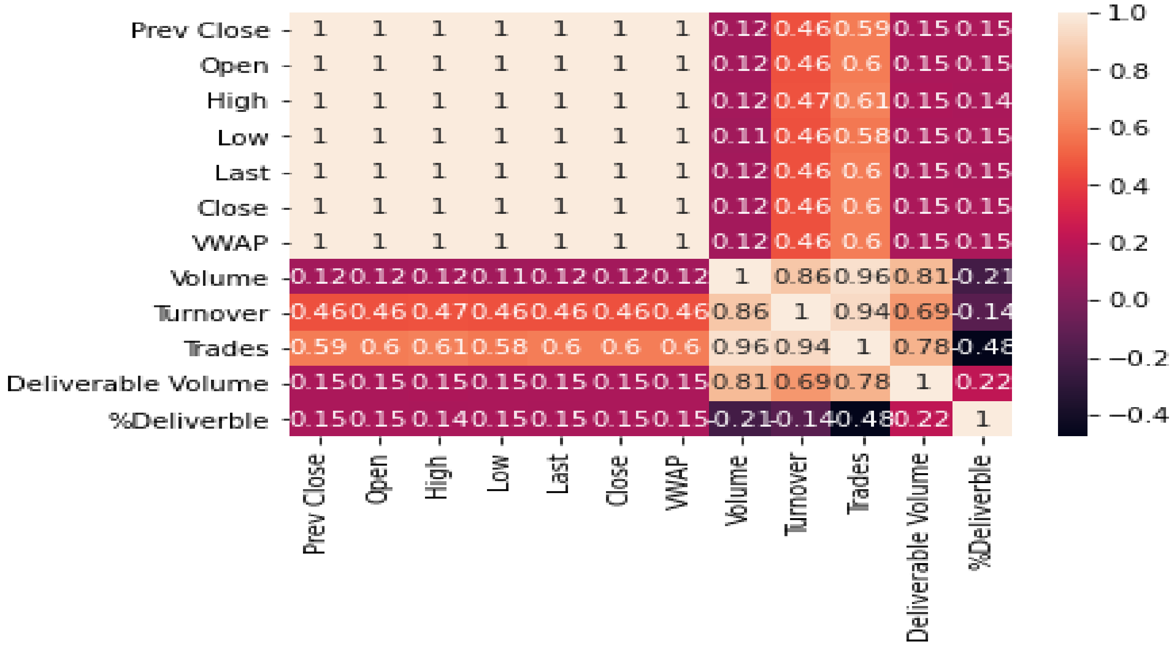


Figure 2: Heatmap reflecting the correlation among attributes

As presented in Figure 2, the heatmap shows the correlation between the attributes that reflect the data exploration statistics. Highest correlation value is 1 and the least correlation value is zero. Thus the heatmap reflects how close or distant an attribute in correlation with other attributes. Just by glance, the data distribution can be understood.

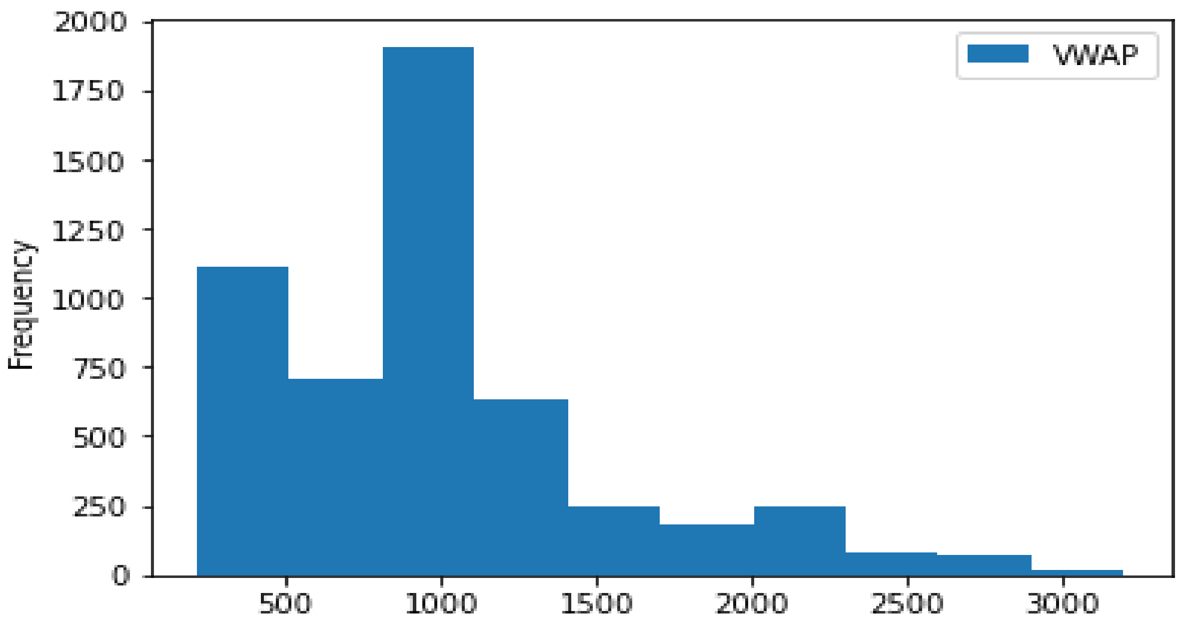


Figure 3: Experimental results showing frequency against VWAP

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006058 |

As presented in Figure 3, the experimental results reveal the benefits of the proposed model in terms of stock market predictions. Volume Weighted Average Price (VWAP) is the measure used for the performance analysis. In fact, it is a technical analysis indicator reflecting performance of the proposed model. The frequency of trades is provided in vertical axis and the VWAP is provided in the horizontal axis.

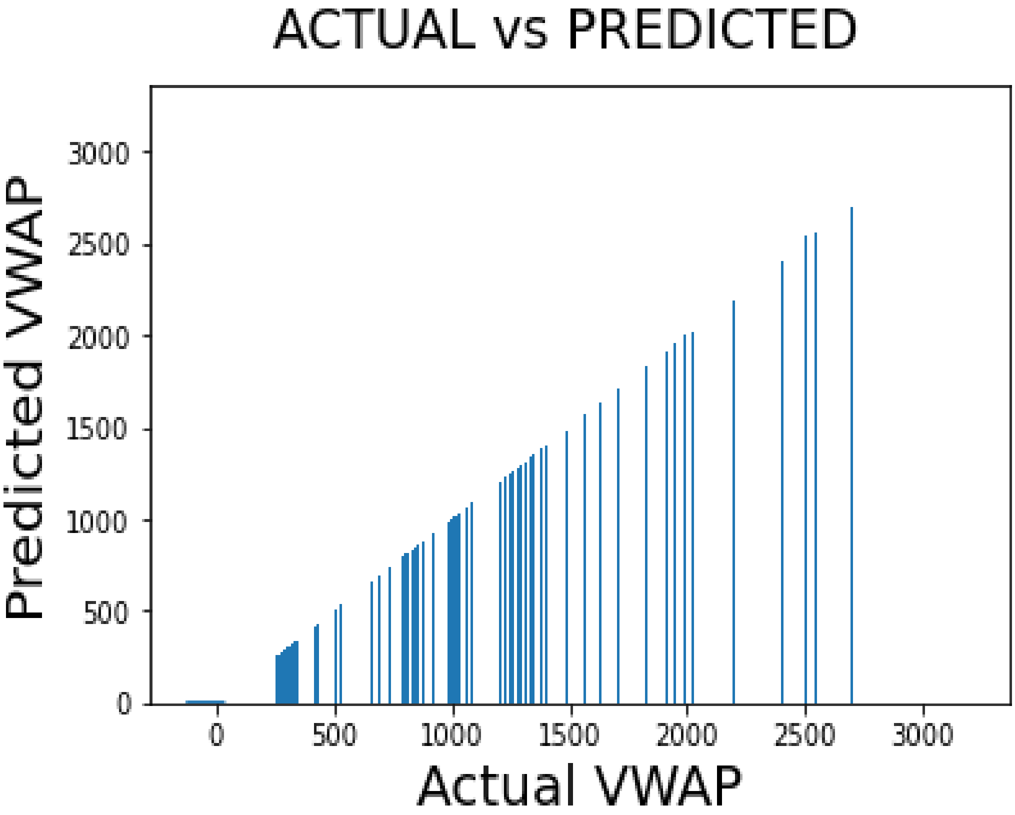


Figure 4: Experimental results showing VWAP forecast

As presented in Figure 4, the experimental results reveal the benefits of the proposed model in terms of stock market predictions. Volume Weighted Average Price (VWAP) is the measure used for the performance analysis. The VWAP is forecasted and compared with the actual VWAP. It is very clear in the graphical representation that there is highest level of accuracy in the prediction process.

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

Table 2: Shows the performance of different ensemble models used for experiments

As presented in Table 2, the accuracy of each ensemble model used to predict stock prices. Each model showed its performance in terms of accuracy in predictions.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006058 |

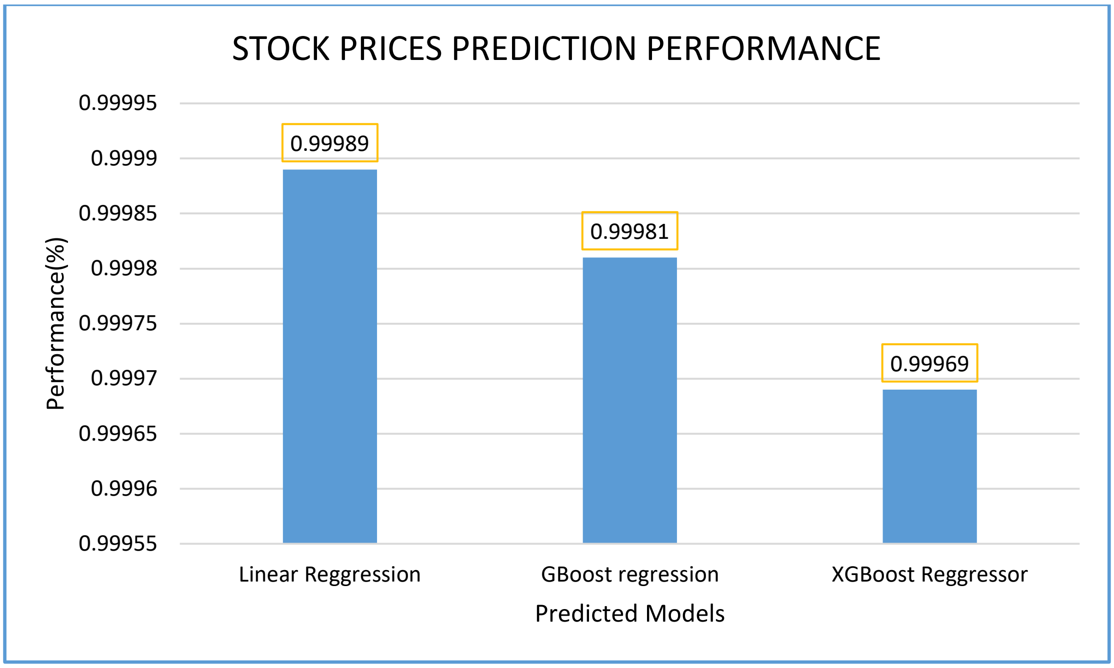


Figure 5: Performance comparison of prediction models

As presented in Figure 5, it is evident that the highest accuracy is achieved by Linear regression. However, the difference among the accuracy of prediction model isvery less. It does mean that all the ensemble models showed 99% accuracy.

V. CONCLUSION AND FUTUER WORK

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]Gholamiangonabadi, Davoud; MohseniTaheri, SeyedDanial; Mohammadi, Afshin; Menhaj, Mohammad Bagher (2014). [IEEE 2014 5th Conference on Thermal Power Plants (CTPP) - Tehran, Iran (2014.6.10-2014.6.11)] 2014 5th Conference on Thermal Power Plants (CTPP) - Investigating the performance of technical indicators in electrical industry in Tehran's Stock Exchange using hybrid methods of SRA, PCA and Neural Networks. , p75–82.

[2] Hyun-jung Kim; Kyung-shik Shin (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. , 7(2), p569–576.

[3]Yakup Kara; MelekAcarBoyacioglu; ÖmerKaanBaykan (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. , 38(5), p5311–5319.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006058 |

[4] Ray Tsaih; Yenshan Hsu; Charles C. Lai (1998). Forecasting S&P 500 stock index futures with a hybrid AI system. , 23(2), p161–174.

[5]SubbaRaoPolamuri;Dr. KudipudiSrinivas;Dr. A. Krishna Mohan; (2021). Multi-Model Generative Adversarial Network Hybrid Prediction Algorithm (MMGAN-HPA) for stock market prices prediction . Journal of King Saud University - Computer and Information Sciences, p1-12.

[6] Mallikarjuna, M.; Rao, R. Prabhakara (2019). Evaluation of forecasting methods from selected stock market returns. Financial Innovation, 5(1), p1-16.

[7] Chung, Hyejung; Shin, Kyung-shik (2018). Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction.Sustainability, 10(10), p1-18.

[8] Amir Hossein Alavi; Amir Hossein Gandomi (2011). Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. , 89(23-24), p2176–2194.

[9] Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process.Expert Systems with Applications, 186, 115716.p1-21.

[10]ShahrokhAsadi; EsmaeilHadavandi; FarhadMehmanpazir; Mohammad MasoudNakhostin (2012). Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. , 35(none), p1-14.

[11]ArashBahrammirzaee (2010). A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. , 19(8), p1165–1195.

[12] Shah, ;Isah, ; Zulkernine, (2019). Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. International Journal of Financial Studies, 7(2), p1-121.

[13]Nti, Isaac Kofi; Adekoya, Adebayo Felix; Weyori, Benjamin Asubam (2019). A systematic review of fundamental and technical analysis of stock market predictions.Artificial Intelligence Review, p1-51.

[14]Karathanasopoulos, Andreas; Theofilatos, KonstantinosAthanasios; Sermpinis, Georgios; Dunis, Christian; Mitra, Sovan; Stasinakis, Charalampos (2015). Stock market prediction using evolutionary support vector machines: an application to the ASE20 index. The European Journal of Finance, p1–19.

[15]RohitChoudhry, and KumkumGarg .(2008). A Hybrid Machine Learning System for Stock Market Forecasting. World Academy of Science, Engineering and Technology.15 (.), p315-318.

[16]Qiu, Mingyue; Song, Yu; Akagi, Fumio (2016). Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. Chaos, Solitons & Fractals, 85, p1–7.

[17] Ticknor, Jonathan L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. Expert Systems with Applications, 40(14), p5501–5506.

[18] Yudong Zhang; Lenan Wu (2009).Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. , 36(5), p8849–8854.

[19] Chen, Yingjun; Hao, Yongtao (2017). A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. Expert Systems with Applications, 80, p340–355.

[20]Fajiang Liu; Jun Wang (2012). Fluctuation prediction of stock market index by Legendre neural network with random time strength function. , 83(none), p12–21.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details