IMPROVED ACCURACY IN PREDICTION OF STOCK EQUITY ANALYSIS AND CLASSIFICATION USING LINEAR REGRESSION AND COMPARED WITH RANDOM FOREST

D. Bharath Kumar[1], S. Magesh Kumar[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The main aim of the research work is to predict the stock equity using the Linear Regression (LR) over the Random Forest (RFA).
Materials and Methods: The two algorithms linear regression and random forest are compared with a sample size = 10. Sample size is calculated using G power software and determined as 10 per group with pretest power 80%, threshold 0.05% and CI 95%.
Results: The analysis of the results shows that the Linear Regression has a high accuracy of (89.36%) in comparison with Random Forest(87.58%). The P value achieved is 0.620 ($p>0.05$), which shows it is a statistically insignificant difference between the study groups.
Conclusion: Prediction in classifying from the results it is concluded that the proposed algorithm Linear Regression will produce better results than the Random forest algorithm.

Keywords: Accuracy, Analysis, Linear Regression, Machine Learning, Novel Stock Equity, Prediction, Random Forest, Stock Price.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and technical Science, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.
[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

1. Introduction

The aim of the work is to predict the Novel Stock equity prediction using Linear Regression (LR) over Random Forest (RFA). The prediction using machine learning has succeeded in LR over Random Forest Algorithm. Stock marketing is the primary source of any industry whether it is private sector or public sector to raise their funds for business expansions and also further growth of the company ("Stock Market Prices Prediction Using Random Forest and Extra Tree Regression" 2019)**.** These questions are the reason why the analysis of stock market price movement has been in the focus of both investors and researchers for quite some time and there should be two common ways of predicting stock equity analysis (Manojlovic and Stajduhar 2015). The study explained that the regression models and the applications can be useful in stock equity prediction (Ravikumar and Saraf 2020). We observe that for both LSTM and autoregressive integrated moving average with variable models a considerable improvement of the prediction of stock price movements. Applications for predicting Novel stock equity include analysis of forecasting financial market prices, including the forecast of stock prices, option pricing, exchange rates, banking and financial crisis and financial status of a company, investing money in stock markets, and improving profit of the company.

In the last five years, Google scholar identified almost 2162 research articles on stock price prediction using machine learning.Stock markets are often interpreted to a particular extent as prediction markets, since they will incorporate and represent the various opinions of investors who disagree on the implications of the available information on past and expected events and trade on their beliefs in order to achieve profits (Lohrmann and Luukka 2019). The goal of predicting the stock price may depend upon the raising or downfall of the market also called as the volatility of the shares. The investment in the stock market is always subjected to market risks ("Stock Price Prediction" 2019). When an RF receives an input of (x), where x is a vector made up of a variety of different evidential features examined for a given training area, the RF builds (N) DTs and averages their results as the final predicted output (Nti, Adekoya, and Weyori 2020). From all these research papers, the best study paper in my opinion is (Manojlovic and Stajduhar 2015); (Ghosh, Neufeld, and Sahoo 2021); (Ravikumar and Saraf 2020)).Our team has extensive knowledge and research experience that has translated into high quality publications(Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

The research gap identified from the existing system Random forest shows poor accuracy. The study is to improve the accuracy of Classification by incorporating LR and comparing performance with RFA. The proposed model improves classifiers to achieve more accuracy for prediction of Novel stock equity analysis.

2. Materials And Methods

This study setting was done in the Data Analytics Laboratory, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The number of required samples in research are two in which group 1 is LR compared with group 2 of Random Forest Algorithm. The samples were taken from the device and iterated 10 times to get desired accuracy with G power 80%, threshold 0.05% and CI 95%. A dataset consisting of a collection of stocks was downloaded from Kaggle website (gracelena 2017)**.**

Linear Regression

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y (output).

Pseudocode for Linear regression

Step1: Import pack.
Step2: Create an input dataset.
Step3: Analyze the size of the taken input data.
Step4: Split the datasets for testing and training the dataset.
Step5: Apply LR algorithm.
Step6: Predict the results.

Random Forest

The Random Forest was utilized for classifying and differentiating input data types. This RFA is widely used in Machine Learning to make predictions. Random Forest is often used in the stock market to predict future profits. It has a big effect on stock price forecasting. So, the program predicts the stock price.

Pseudocode for Random Forest

Step1: Import packages.
Step2: Create an input dataset.
Step3: Analyze the size of the taken input data.
Step4: Split the datasets for testing and training the dataset.
Step5: Apply Random Forest algorithm.
Step6: Predict the results

Recall that the testing setup includes both hardware and software configuration choices. The laptop has an Intel Core i5 8th generation CPU with 12GB of RAM, an x86-based processor, a 64-bit operating system, and a hard drive. Currently, the software runs on Windows 10 and is programmed in Python. Once the program is finished, the accuracy value will appear. Procedure: Wi-Fi laptop connected. Chrome to Google Collaboratory search Write the code in Python. Run the code. To save the file, upload it to the disc, and create a folder for it. Log in using the ID from the message. Run the code to output the accuracy and graph.

Statistical Analysis

SPSS is a software tool used for statistics analysis. The proposed system utilized 10 iterations for each group with predicted accuracy noted and analyzed. Independent samples t-test was done to obtain significance between two groups. Stock price opening and closing price parameters are independent variables and stock price prediction is dependent variable. Structure of this paper we describe the historical data and input parameters used for building the proposed models. Then details on the machine learning model used in the process of presenting the results of empirical evaluation (Parmar et al. 2018)**.**

3. Results

Table 1 shows the accuracy value of iteration of LR and RFA.

Table 2 represents the Group statistics results which depicts LR with mean accuracy of 89.36%, and standard deviation is 3.27. Random Forest has a mean accuracy of 87.58% and standard deviation is 2.99. Proposed LR algorithm provides better performance compared to the Random Forest algorithm.

Table 3 shows the independent samples T-test value for LR and Random Forest with Mean difference as 1.78, std Error Difference as 1.40. Significance value is observed as 0.620 ($p>0.05$), which shows that two groups are statistically insignificant.

Figure 1 shows the bar graph comparison of mean of accuracy on LR and Random Forest algorithm. Mean accuracy of LR is 89.36% and Random Forest is 87.58%.

4. Discussion

In this study, predicting Novel stock equity in the stock market using the LR algorithm has significantly higher accuracy, approximately (89.36%) in comparison to Random Forest (87.58%). LR appears to produce more consistent results with minimal standard deviation.

We plan to maximize the prediction of stocks by collecting and reviewing data with the help of Twitter API using Random Forest Regressor. There are a variety of variables affecting stock markets, the foremost important of which are historical records (Padmanayana et al. 2021). Random forest algorithm constructs the multiple decision tree at training time and outputs the category that is the mode of the category for the classification task (Narkar 2019). Each factor was studied independently to seek out its association with market performance. The results suggested that market behavior of the market is often predicted using machine learning techniques (Raza 2017). Now we have the idea that if we catch the timing signal, the movement of the Novel stock equity has been in range then we will invest at that time during a random forest, each node is split using the simplest among a subset of predictors randomly chosen at that node (Werawithayaset and Tritilanunt 2019). The aim of this paper is to predict clean energy stock equity direction using random forests. Directional stock price forecasts are constructed from within the future; a somewhat counterintuitive strategy seems to perform alright compared to several other classifiers (Sadorsky 2021).

The limitation of this research is that complexity of stock equity data, development of efficient models for predicting is extremely difficult. Stock equity is user privacy that users may hesitate to share personal information with the software. During this model it is not able to consider all given feature variable parameters for training. The longer term scope of proposed work is going to be prediction of stock equity supported classification using class labels for lesser time complexity.

5. Conclusion

In this study, prediction of Novel stock equity analysis using Linear Regression algorithm provides better accuracy than Random Forest algorithm. Accuracy of linear regression is (89.36) and the Random Forest is (87.58). By finding the accuracy of both algorithms linear regression has higher accuracy.

Declaration

Conflict of Interests

No conflict of interests in this manuscript

Authors Contribution

Author DBK was involved in data collection, data analysis, and manuscript writing. Author SMK was

involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgement
The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary Infrastructure to carry out this work successfully.

Funding
We thank the following organizations for providing financial support that enabled us to complete the study.

1. Edify Technology Services.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." *Chemosphere*. https://doi.org/10.1016/j.chemosphere.2022.134596.

Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." *Bioresource Technology*. https://doi.org/10.1016/j.biortech.2022.127234.

Ghosh, Pushpendu, Ariel Neufeld, and Jajati Keshari Sahoo. 2021. "Forecasting Directional Movements of Stock Prices for Intraday Trading Using LSTM and Random Forests." *Finance Research Letters*. https://doi.org/10.1016/j.frl.2021.102280.

gracelena. 2017. "Uniqlo Stock Price Prediction." Kaggle. August 27, 2017. https://kaggle.com/gracelena/uniqlo-stock-price-prediction.

Karpagam, M., R. Beaulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." *Soft Computing*. https://doi.org/10.1007/s00500-022-06931-1.

Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation." *Journal of Thermal Analysis and Calorimetry*. https://doi.org/10.1007/s10973-022-11298-4.

Lohrmann, Christoph, and Pasi Luukka. 2019. "Classification of Intraday S&P500 Returns with a Random Forest." *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2018.08.004.

Manojlovic, T., and I. Stajduhar. 2015. "Predicting Stock Market Trends Using Random Forests: A Sample of the Zagreb Stock Exchange." *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. https://doi.org/10.1109/mipro.2015.7160456.

Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. "Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm." *Computational Intelligence and Neuroscience* 2022 (April): 6461690.

Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands." *Silicon*. https://doi.org/10.1007/s12633-022-01825-1.

Narkar, Aditya Vijay. 2019. *Stock Price Prediction Using Feature Engineering and Machine Learning Techniques*.

Nti, Isaac Kofi, Adebayo Felix Adekoya, and Benjamin Asubam Weyori. 2020. "Efficient Stock-Market Prediction Using Ensemble Support Vector Machine." *Open Computer Science*. https://doi.org/10.1515/comp-2020-0199.

Padmanayana, Padmanayana, Varsha, and K. Bhavya. 2021. "Stock Market Prediction Using Twitter Sentiment Analysis." *International Journal of Scientific Research in Science and Technology*. https://doi.org/10.32628/cseit217475.

Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification

Using Stand-Alone and Hybrid Energy Technologies." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102155.

Parmar, Ishita, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, and Lokesh Chouhan. 2018. "Stock Market Prediction Using Machine Learning." In *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*. IEEE. https://doi.org/10.1109/icsccc.2018.8703332.

Ravikumar, Srinath, and Prasad Saraf. 2020. "Prediction of Stock Prices Using Machine Learning (Regression, Classification) Algorithms." *2020 International Conference for Emerging Technology (INCET)*. https://doi.org/10.1109/incet49848.2020.9154061.

Raza, Kamran. 2017. "Prediction of Stock Market Performance by Using Machine Learning Techniques." *2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT)*. https://doi.org/10.1109/icieect.2017.7916583.

Sadorsky, Perry. 2021. "A Random Forests Approach to Predicting Clean Energy Stock Prices." *Journal of Risk and Financial Management*. https://doi.org/10.3390/jrfm14020048.

"Stock Market Prices Prediction Using Random Forest and Extra Tree Regression." 2019. *International Journal of Recent Technology and Engineering* 8 (3): 1224–28.

"Stock Price Prediction." 2019. *Special Issue Report. International Whaling Commission* 9 (2S): 425–29.

Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. "Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends." *Energy*. https://doi.org/10.1016/j.energy.2022.123806.

Werawithayaset, Pawee, and Suratose Tritilanunt. 2019. "Stock Closing Price Prediction Using Machine Learning." *2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE)*. https://doi.org/10.1109/ictke47035.2019.8966836.

Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. "Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture." *Chemosphere* 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review." *Chemosphere* 299 (July): 134390.

Tables And Figures

Table 1. Accuracy Values for LR and RFA

S.NO	LR	RFA
1	94.40	90.78
2	92.45	90.56
3	89.23	87.39
4	87.12	86.89
5	90.63	89.16
6	91.79	90.72
7	85.25	83.49
8	84.59	82.28
9	91.25	88.61

10	86.92	85.95

Table 2. Group Statistics Results-LR has an mean accuracy (89.36%), std.deviation (3.27), whereas for RFA has mean accuracy (87.58%), std.deviation (2.99).

Group Statistics					
Accuracy	**Groups**	**N**	**Mean**	**Std deviation**	**Std. Error Mean**
	LR	**10**	89.363	3.27741	1.03641
	RFA	**10**	87.583	2.99118	.94589

Table 3. Independent Samples T-test - LR shows significance value achieved is p=0.620 (p>0.05), which shows that two groups are statistically insignificant.

Accuracy	**Independent Samples Test**								
	Levene's Test for Equality of Variances		**T-test for Equality of Means**						
	F	**Sig**	**t**	**df**	**Sig(2-tailed)**	**Mean Difference**	**Std.Error Difference**	**95% Confidence Interval of the Difference**	
								Lower	**Upper**
Equal variances assumed	.255	.620	1.26	18	.221	1.78000	1.40316	-1.1679	4.72793
Equal variances not assumed			1.26	17.85	.221	1.78000	1.40316	-1.1696	4.72968

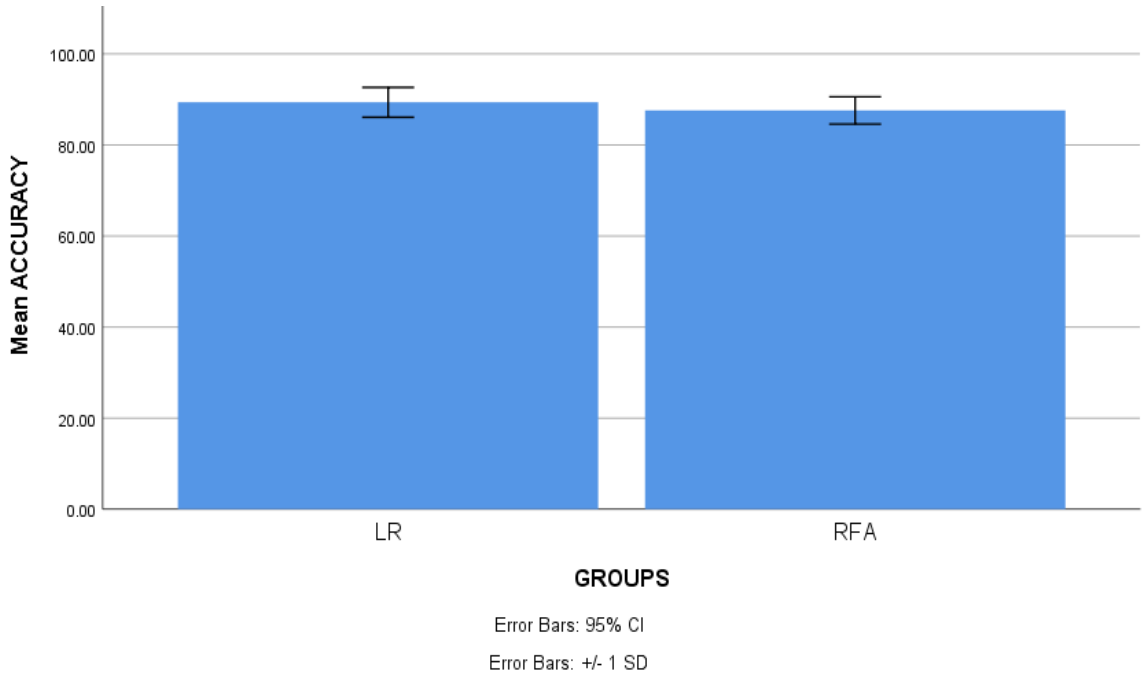

Fig. 1. Bar Graph Comparison on mean accuracy of LR (89.36%) and RFA (87.58%). X-axis: LR, RFA, Y-axis: Mean Accuracy with ±1 SD.