



IMPROVED ACCURACY IN PREDICTION OF STOCK EQUITY ANALYSIS AND CLASSIFICATION USING LINEAR REGRESSION AND COMPARED WITH DECISION TREE.

D. Bharath Kumar¹, 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 Decision Tree(DTA).

Materials and Methods: The two algorithms linear regression and decision tree 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 (94.38%) in comparison with the Decision Tree (89.47). Attained Significance Accuracy value is 0.894 ($p > 0.05$). There is a statistically insignificant difference between the study groups with these algorithms.

Conclusion: Prediction in classifying from the results it is concluded that the proposed algorithm Linear Regression will produce better results than the Decision tree algorithm.

Keywords: Linear Regression, Machine Learning, Novel Stock Equity, Prediction, Stock Price, Decision Tree.

¹Research 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 work is about Novel Stock equity prediction using Linear Regression(LR) over Decision Tree(DT). The prediction using machine learning has succeeded in LR over Decision Tree Algorithm. This information, such as news about the company can influence people's decision whether or not they will buy the company's stock. More people buy the company's stock, the price is getting higher (Sable, Goel, and Chatterjee 2019). There is a misconception about the stock market i.e. buying and selling of shares is an act of gambling. There is a huge amount of data generated by stock markets forcing the researchers to apply data mining to make investment decisions. There is a huge amount of data generated by stock markets forcing the researchers(Sharma, Bhuriya, and Singh 2017). However, it was observed that solely historical data does not give the predictions accurately. Various new disciplines came into existence which provided better prediction models. In context with stock market prediction, many researchers have been able to devise models for stock market prediction (Singh et al. 2019). The effort to find out the upcoming stock market value is recognized as a Novel stock equity prediction. The forecast is anticipated to be efficient, accurate and robust. Applications (Kadu and Bamnote 2021) for predicting 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 21600 research articles on stock price prediction using machine learning. Stock equity prediction is full of challenges, and data scientists usually confront some problems when they try to develop a predictive model (Nabipour et al. 2020). Because many brains will try to interpret the pattern and make a guess and this such activity has proven to be a lot more successful in practice than it seems in theory. Traders and buyers continue selling these shares at their own price but the company only gets to keep the money made during the IPO (Pahwa and Agarwal 2019). One decision in the Stock Market can make a huge impact on an investor's life. The Novel stock equity is a complex system and often covered in mystery and it is very difficult to analyze all the impacting factors before making a decision (Raza 2017). This study uses stock markets to predict the analysis respectively with historical data, futures, and options as data sets to predict stock prices in these two markets (Wu et al. 2021). From all these research papers, the best study paper in my opinion

is (Sharma, Bhuriya, and Singh 2017);(Sable, Goel, and Chatterjee 2019);(Kadu and Bamnote 2021). 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 Decision Tree shows poor accuracy. The study is to improve the accuracy of Classification by incorporating LR and comparing performance with Decision Tree. The proposed model improves classifiers to achieve more accuracy for prediction of 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 Decision Tree 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.

Decision Tree

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

Pseudocode for Decision Tree

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 Decision Tree 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. Novel Stock equity opening and closing price parameters are Independent variables and stock price prediction is Dependent variable (Magee 1964).

3. Results

Table 1 shows the accuracy value of iteration of LR and Decision Tree. Table 2 represents the Group statistics results which depicts LR with mean accuracy of 91.81%, and standard deviation is 2.27. Decision Tree has a mean accuracy of 88.75% and standard deviation is 2.43. Proposed LR algorithm provides better performance compared to the Decision Tree algorithm. Table 3 shows the independent samples T-test value for LR and Decision Tree with Mean difference as 3.06, std Error Difference as 1.05. Significance value is observed as 0.89 ($p > 0.05$), which shows it is a statistically insignificant difference between the study groups.

Figure 1 shows the bar graph comparison of mean of accuracy on LR and Decision Tree algorithm. Mean accuracy of LR is 91.81% and Decision Tree is 88.75%.

4. Discussion

In this study, predicting Novel stock equity in the stock market using the LR algorithm has significantly higher accuracy, approximately

(91.81%) in comparison to Decision Tree (88.75%). LR appears to produce more consistent results with minimal standard deviation.

The algorithms can be effectively self-learning, and can tackle the predicting task of price fluctuations in order to improve trading strategies. The prediction results not only are affected by the representation of the input data but also depend on the prediction method (Chen 2020). The prediction results not only are affected by the representation of the input data but also depend on the prediction method. Researchers are conducting research on stock market forecasts from a variety of fields, including computer science and business. This type of prediction is even more appealing when it involves money and risks such as Novel Stock Market speculation (Lv et al. 2021). This exact phenomenon is the reason for the fear people have in investing in stock markets and the reason for the fall and rise of stock prices in a nutshell. The prediction results not only are affected by the representation of the input data but also depend on the prediction method (Qiu, Wang, and Zhou 2020). Results indicated that the SVM outperformed other predictive models. The price fluctuation by a developed Legendre neural network was forecasted by Liu et al by assuming investors' positions and their decisions by analyzing the prior data on the stock values (Lee et al. 2022).

The limitation of this research is that complexity of stock equity data, development of efficient models for predicting is extremely difficult. Stock prices are 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 Novel 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 Decision Tree algorithm. Accuracy of linear regression is 91.81% and the Decision Tree is 88.75%. By finding the accuracy of both algorithms linear regression has higher accuracy.

Declaration

Conflict of Interest

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>.
- Chen, Lumeng. 2020. *Stock Price Prediction Using Adaptive Time Series Forecasting and Machine Learning Algorithms*.
- 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>.
- gracelena. 2017. "Uniqlo Stock Price Prediction." Kaggle. August 27, 2017. <https://kaggle.com/gracelena/uniqlo-stock-price-prediction>.
- Kadu, Parag P., and G. R. Bamnote. 2021. "Comparative Study of Stock Price Prediction Using Machine Learning." In *2021 6th International Conference on Communication and Electronics Systems (ICCES)*. IEEE. <https://doi.org/10.1109/icc51350.2021.9489170>.
- Karpagam, M., R. Beulah 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>.
- Lee, Ming-Che, Jia-Wei Chang, Sheng-Cheng Yeh, Tsorng-Lin Chia, Jie-Shan Liao, and Xu-Ming Chen. 2022. "Applying Attention-Based BiLSTM and Technical Indicators in the Design and Performance Analysis of Stock Trading Strategies." *Neural Computing & Applications*, January, 1–13.
- Lv, Jiehua, Chao Wang, Wei Gao, and Qiumin Zhao. 2021. "An Economic Forecasting Method Based on the LightGBM-Optimized LSTM and Time-Series Model." *Computational Intelligence and Neuroscience* 2021 (September): 8128879.
- Magee, John F. 1964. *Decision Trees for Decision Making*.
- Nabipour, Mojtaba, Pooyan Nayyeri, Hamed Jabani, Shahab, and Amir Mosavi. 2020. "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms via Continuous and Binary Data; A Comparative Analysis." *IEEE Access: Practical Innovations, Open Solutions* 8: 150199–212.
- 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 Babu, 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>.
- Pahwa, Kunal, and Neha Agarwal. 2019. "Stock Market Analysis Using Supervised Machine Learning." In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*. IEEE.

- <https://doi.org/10.1109/comitcon.2019.8862225>.
- Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabakaran, 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>.
- Qiu, Jiayu, Bin Wang, and Changjun Zhou. 2020. "Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism." *PloS One* 15 (1): e0227222.
- Raza, Kamran. 2017. "Prediction of Stock Market Performance by Using Machine Learning Techniques." In *2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT)*. IEEE. <https://doi.org/10.1109/icieect.2017.7916583>.
- Sable, Rachna, Shivani Goel, and Pradeep Chatterjee. 2019. "Empirical Study on Stock Market Prediction Using Machine Learning." In *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*. IEEE. <https://doi.org/10.1109/icac347590.2019.9036786>.
- Sharma, Ashish, Dinesh Bhuriya, and Upendra Singh. 2017. "Survey of Stock Market Prediction Using Machine Learning Approach." In *2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA)*. IEEE. <https://doi.org/10.1109/iceca.2017.8212715>.
- Singh, Sukhman, Tarun Kumar Madan, Jitendra Kumar, and Ashutosh Kumar Singh. 2019. "Stock Market Forecasting Using Machine Learning: Today and Tomorrow." In *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*. IEEE. <https://doi.org/10.1109/icicict46008.2019.8993160>.
- 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>.
- 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.
- Wu, Jimmy Ming-Tai, Lingyun Sun, Gautam Srivastava, and Jerry Chun-Wei Lin. 2021. "A Long Short-Term Memory Network Stock Price Prediction with Leading Indicators." *Big Data* 9 (5): 343–57.
- 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 Decision Tree

.NO	LR	Decision Tree
1	94.38	89.47
2	92.45	90.12
3	87.32	86.03
4	93.50	91.92
5	93.30	89.19
6	94.30	92.08
7	91.60	87.40

8	89.30	84.29
9	91.30	87.99
10	90.66	89.01

Table 2. Group Statistics Results-LR has an mean accuracy (91.81%), std.deviation (2.27), whereas for Decision Tree has mean accuracy (88.75%), std.deviation (2.43).

Group Statistics					
Accuracy	Groups	N	Mean	Std deviation	Std. Error Mean
	LR	10	91.8110	2.27004	.71785
	Decision Tree	10	88.7500	2.43777	.77089

Table 3. Independent Samples T-test - LR shows significance value achieved is $p=0.894$ ($p>0.05$), 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	.018	.894	2.906	18	.009	3.06100	1.05337	.84796	5.27404
Equal variances not assumed			2.906	17.909	.009	3.06100	1.05337	.84716	5.27484

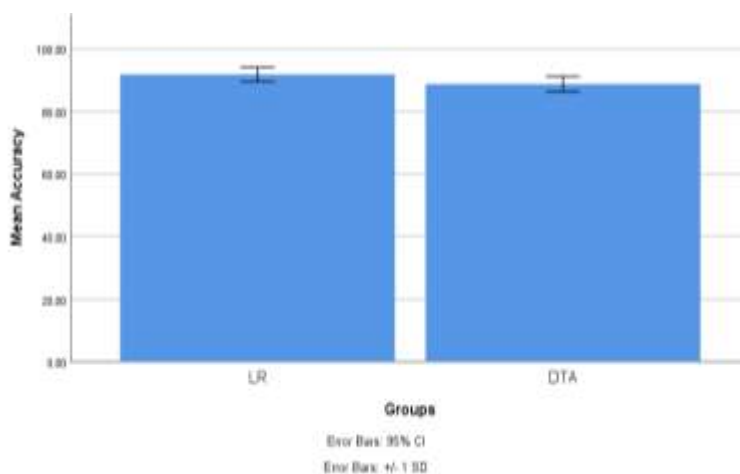


Fig. 1. Bar Graph Comparison on mean accuracy of LR (91.81%) and Decision Tree (88.75%). X-axis: LR, RFA, Y-axis: Mean Accuracy with ± 1 SD and 95% CI.