Improved Accuracy in Prediction of Stock Equity Analysis and Classification using Linear Regression and compared with K Nearest Neighbor

Section A-Research paper

IMPROVED ACCURACY IN PREDICTION OF STOCK EQUITY ANALYSIS AND CLASSIFICATION USING LINEAR REGRESSION AND COMPARED WITH K NEAREST NEIGHBOR

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

Abstract

Aim: The main aim of the research work is to predict the stock equity using the Linear Regression (LR) over the K Nearest Neighbor (KNN).

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 (86.19%). Attained Significance Accuracy value is 0.591 ($p<0.05$). 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: K Nearest Neighbor, Linear Regression, Machine Learning, Novel Stock Equity, Prediction, Stock Market.

[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.

Eur. Chem. Bull. 2023, 12 (S1), 4686– 4691

4686

Improved Accuracy in Prediction of Stock Equity Analysis and Classification using Linear Regression and compared with K Nearest Neighbor

Section A-Research paper

1. Introduction

The work is about Novel Stock equity prediction using Linear Regression (LR) over K Nearest Neighbour (KNN). The prediction using machine learning has succeeded in LR over K Nearest Neighbour Algorithm. Stock price forecasting needs a huge amount of information in the process of modeling and forecasting, which puts forward higher requirements for the ability of the algorithm to process massive data (Lv et al. 2021). However, these two kinds of information are too complex and unstable to gather. In the present scenario of the financial market world, especially in the stock market, forecasting the trend or the price of stocks using machine learning techniques and artificial neural networks are the most attractive issues to be investigated (Bosco and Khan 2018). Stock investors and financial planners use various time series methods for market prediction techniques. There is a higher possibility that a stock price will continue to rise if the sentiment is positive, and the Novel Stock equity will fall if it is malicious (Mehta, Pandya, and Kotecha 2021). The algorithms can be effectively self-learning, and can tackle the predicting task of price fluctuations in order to improve trading strategies Moreover, using only prominent features and identifying them as input data instead of all features can noticeably develop the accuracy of the prediction models (Lee et al. 2022). 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 1563 research articles on stock price prediction using machine learning. The study explained the regression models and how their application can be useful in price prediction. This study gave us an idea about how choosing appropriate factors affecting stock price as variables can give some predictions (Usmani and Shamsi 2021). Sentiment analysis is a widely used approach to infer emotion from textual data that represents subjective information. It is a major area of interest in today's text analytics (Lin et al. 2021). The classification of stock market based on stock performance, stock equity opening and closing price parameters were included in stock price prediction (Misra, Yadav, and Kaur 2018). As the market's ability to forecast stock prices declines, investor risk increases. Only those who can predict the Novel stock equity before investing may avoid losses (Obthong et al. 2020). From all these research papers, the best study paper in my opinion is (Usmani and Shamsi 2021); (Mehta, Pandya, and Kotecha 2021); (Lv et al. 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 K Nearest Neighbour shows poor accuracy. The study is to improve the accuracy of Classification by incorporating LR and comparing performance with K Nearest Neighbour. 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 K Nearest Neighbour 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**.**

K Nearest Neighbor
The K Nearest Neighbour was utilized for classifying and differentiating input data types. This KNN is widely used in Machine Learning to make predictions. K Nearest Neighbour is often used in the Novel stock market to predict future profits. It has a big effect on stock price

Eur. Chem. Bull. 2023, 12 (S1), 4686– 4691

4687

Improved Accuracy in Prediction of Stock Equity Analysis and Classification using Linear Regression and compared with K Nearest Neighbor

Section A-Research paper

forecasting. So, the program predicts the stock price.

Pseudocode for K Nearest Neighbour
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 K Nearest Neighbour 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 (Chen and Shah 2018) opening and closing price parameters are Independent variables and stock price prediction is Dependent variable.

3. Results

Table 1 shows the accuracy value of iteration of LR and K Nearest Neighbour. Table 2 represents the Group statistics results which depicts LR with mean accuracy of 91.81%, and standard deviation is 2.27. K Nearest Neighbour has a mean accuracy of 87.86% and standard deviation is 2.47. Proposed LR algorithm provides better performance compared to the K Nearest Neighbour algorithm. Table 3 shows the independent samples T-test value for LR and KNN with Mean difference as 3.94, std Error Difference as 1.06. Significance value is observed as $0.591(p>0.05)$, 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 K Nearest Neighbour algorithm. Mean accuracy of LR is 91.81% and K Nearest Neighbor is 87.86 %.

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 K Nearest Neighbour (87.86 %). LR appears to produce more consistent results with minimal standard deviation.

The accuracy of your classifier is directly proportional to the amount of data provided to the classifier and the attributes selected with the important decision parameter as the volume traded has the most direct impact on future stock price than any other feature (Misra, Yadav, and Kaur 2018). The empirical results show that the proposed techniques can predict the up and down on a Novel stock equity after the news is announced or released. The proposed method presented in the study is straightforward, simple and valuable for the short-term investors (Jarrah and Salim 2016). However, the time-series of the stock is not a function that can map easily. It tends to best depict as a random walk, where it makes the prediction and feature engineering much harder (Herwartz and Kholodilin 2014). A comparative assessment of these approaches is timely, given the scope of artificial neural networks for the prediction of stock markets to identify stock trends from large amounts of data, but there is so far, no such algorithm or model that could consistently predict stock prices (Reddy and Usha 2019). The stock market crisis can emerge due to variations in global and local market economic policy and macroeconomic data and these studies considered the less than one percentile of stock returns as a stock crisis point has not been considered a fundamental analysis to identify stock quality because stock technical analysis performs better on the strong fundamental stock (Naik and Mohan 2021).

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 stock equity supported classification using class labels for lesser time complexity.

5. Conclusion

In this study, prediction of stock equity analysis using Linear Regression algorithm provides better accuracy than K Nearest Neighbour algorithm. Accuracy of linear regression is 91.81% and the K Nearest Neighbour is 87.86%. By finding the

Eur. Chem. Bull. 2023, 12 (S1), 4686– 4691

4688

Improved Accuracy in Prediction of Stock Equity Analysis and Classification using Linear Regression and compared with K Nearest Neighbor

Section A-Research paper

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.

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.

Bosco, Joish, and Fateh Khan. 2018. *Stock Market Prediction and Efficiency Analysis Using Recurrent Neural Network*. GRIN Verlag.

Chen, George H., and Devavrat Shah. 2018. *Explaining the Success of Nearest Neighbor Methods in Prediction*. Foundations and Trends (R) in Machine Learning.

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.

Herwartz, Helmut, and Konstantin A. Kholodilin. 2014. "In-Sample and Out-of-Sample Prediction of Stock Market Bubbles: Cross-Sectional Evidence." *Journal of Forecasting*. https://doi.org/10.1002/for.2269.

Jarrah, Mu'tasem, and Naomie Salim. 2016. "Stock Market Prediction Based on Term Frequency-Inverse Document Frequency." *Journal of Economics, Business and Management*. https://doi.org/10.7763/joebm.2016.v4.388.

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.

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.

Lin, Yaohu, Shancun Liu, Haijun Yang, Harris Wu, and Bingbing Jiang. 2021. "Improving Stock Trading Decisions Based on Pattern Recognition Using Machine Learning Technology." *PloS One* 16 (8): e0255558.

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.

Mehta, Pooja, Sharnil Pandya, and Ketan Kotecha. 2021. "Harvesting Social Media Sentiment Analysis to Enhance Stock Market Prediction Using Deep Learning." *PeerJ. Computer Science* 7 (April): e476.

Misra, Meghna, Ajay Prakash Yadav, and Harkiran Kaur. 2018. "Stock Market Prediction Using Machine Learning Algorithms: A Classification Study." *2018 International*

Eur. Chem. Bull. 2023, 12 (S1), 4686– 4691

4689

Improved Accuracy in Prediction of Stock Equity Analysis and Classification using Linear Regression and compared with K Nearest Neighbor

Section A-Research paper

Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE). https://doi.org/10.1109/icrieece44171.2018.9009178.

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.

Naik, Nagaraj, and Biju R. Mohan. 2021. "Novel Stock Crisis Prediction Technique—A Study on Indian Stock Market." *IEEE Access.* https://doi.org/10.1109/access.2021.3088999.

Obthong, Mehtabhorn, Nongnuch Tantisantiwong, Watthanasak Jeamwatthanachai, and Gary Wills. 2020. "A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques." *Proceedings of the 2nd International Conference on Finance, Economics, Management and IT Business.* https://doi.org/10.5220/0009340700630071.

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.

Reddy, Bhanuteja, and J. C. Usha. 2019. "Prediction of Stock Market Using Stochastic Neural Networks." *International Journal of Innovative Research in Computer Science & Technology.* https://doi.org/10.21276/ijircst.2019.7.5.1.

Usmani, Shazia, and Jawwad A. Shamsi. 2021. "News Sensitive Stock Market Prediction: Literature Review and Suggestions." *PeerJ. Computer Science* 7 (May): e490.

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.

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 K Nearest Neighbor

S.NO	LR	K Nearest Neighbor
1	94.38	86.19
2	92.45	89.28
3	87.32	83.49
4	93.50	90.14
5	93.30	89.70
6	94.30	91.34
7	91.60	87.18

Eur. Chem. Bull. 2023, 12 (S1), 4686– 4691

4690

*Improved Accuracy in Prediction of Stock Equity
Analysis and Classification using Linear Regression
and compared with K Nearest Neighbor*

Section A-Research paper

8	89.30	86.92
9	91.30	85.16
10	90.66	89.22

Table 2. Group Statistics Results-LR has an mean accuracy (91.81%), std.deviation (2.27), whereas KNN has mean accuracy (87.86%), std.deviation (2.47).

Group Statistics					
	Groups	**N**	**Mean**	**Std deviation**	**Std. Error Mean**
Accuracy	**LR**	**10**	91.81	2.27	.71
	K Nearest Neighbour	**10**	87.86	2.47	.78

Table 3. Independent Samples T-test - LR shows significance value achieved is p=0.591 (p>0.05), which shows that two groups are statistically insignificant.

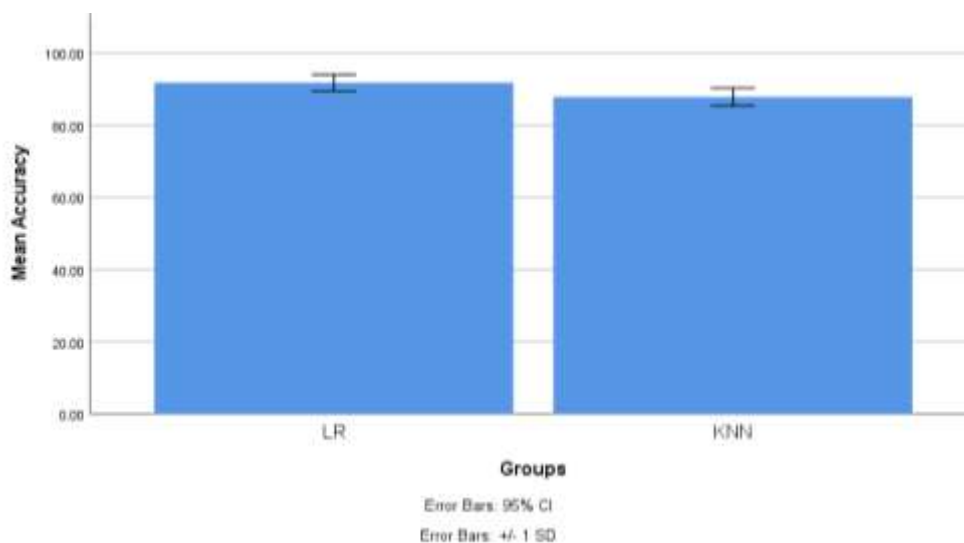| Accuracy	Independent Samples Test								
	Levene's Test for Equality of Variances		**T-test for Equality of Means**						
	F	**Sig**	**t**	**df**	**Sig(2-tailed)**	**Mean Difference**	**Std.Error Difference**	**95% Confidence Interval of the Difference**	
								Lower	**Upper**
Equal variances assumed	.299	.591	3.719	18	.002	3.94900	1.06176	1.71833	6.17967
Equal variances not assumed			3.719	17.868	.002	3.94900	1.06176	1.71715	6.18085

Fig. 1. Bar Graph Comparison on mean accuracy of LR (91.81%) and KNN (87.86%). X-axis: LR, KNN, Y-axis: Mean Accuracy with ± 1 SD and CI - 95.

Eur. Chem. Bull. 2023, 12 (S1), 4686– 4691

4691