STOCK MARKET PREDICTION USING LINEAR REGRESSION ALGORITHM BY COMPARING WITH ARTIFICIAL NEURAL NETWORK ALGORITHM TO IMPROVE ACCURACY

Myneni. Keerthi Pranav[1], S. Ashok Kumar[2*]

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

Abstract

Aim: The main aim of this research study is to get the error free and accurate stock prices using novel approaches of linear regression and comparing it with Artificial neural networks.
Materials and Methods: In this research study Linear regression algorithm is being compared with Artificial neural network algorithm for estimating Stock market values and error free values, which helps the users for availing the profits by estimating the previous sales by analyzing it through defining the set of modules . Sample size is determined by using the G power calculator and found to be 158 per group. Total of 2 groups are used.
Results: Based on the analysis done Linear regression algorithm method has an accuracy of 85% and Artificial neural network has 73% and the significance value achieved is 1.000 ($p>0.05$). It shows that two groups are statistically insignificant.
Conclusion: In this research study, the proposed Linear regression algorithm has shown the highly predicted values when compared to the Artificial neural network.

Keywords: Novel approach, Linear Regression, Artificial neural network, Accuracy, analysis, Stock Market

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu India.
[2*]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu, India.

1. Introduction

Stock market prediction has been one of the trendiest and fast-growing businesses in the present world. The main objective of the research study is to provide future accurate prices of stocks for a definite period of time(Syed 2021). Errors in the predicted prices lead to the loss in the investments of the investors which makes them lose their financial income. Dataset is the novel approach prerequisite required in predicting the stock market values(Alfonso et al. 2021). Data mining and machine learning techniques are the most used techniques for analyzing the stock prices which provide the accurate values(Shaikh and Shinde, n.d.). Data mining techniques consist of various models such as linear regression model, KNN model, ANN model, and apriori algorithm(Upadhyay et al. 2016).

Stock market predictions using linear regression are implemented for prediction to benefit stock investors: nearly 150 in IEEE explorer and 240 in Google scholar(Erb, n.d.). The articles that are mostly cited based on the most viewed are one of them depending on the study of stock market prediction. The current trend of fast growing business mainly consists of the stock market(Vohra and Tanna 2021). The previous research studies state that analyzing stock prices is a financially risky and a challenging task as this deals with the financial data and financial markets(Poitras, n.d.). Machine learning and data mining techniques are mainly suitable for forecasting the data such as stock market prediction and weather forecasting using the novel approaches(Nayak, Pai, and Pai 2016)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 limitations of stock investors is to improve the efficiency of the outcome of the predicted values of stock market prices. In this research it is observed that Linear Regression has brought out the outcome of more accuracy than Artificial neural network algorithm(Jain, Arya, and Singh 2020). The main aim of research is to predict the efficient prices of stock using Linear regression model for better outcomes using different novel approaches(Wang and Guo 2020).

2. Materials and Methods

The research work is carried out in DBMS Laboratory, Department of Computer Science Engineering, Saveetha School of Engineering. In this study 2 sample groups were taken.The group 1 was Linear Regression model and group 2 was Artificial neural network algorithm. Sample size is calculated using Gpower software considering the pretest power to be 80% and CI of 80% (Matloff 2017). The work has been carried out with 3500 records which were taken from a kaggle data set. The accuracy in predicting loyalties was initiated by two different groups.Totally 10 iterations were analyzed and performed on each group to accomplish maximum accuracy. Dataset contains 3000 instances and 7 attributes named Date, High, Low, Close, Open, Volume, OpenInt.Here data is from kaggle website (Kaggle: Your Machine Learning and Data Science Community).

Linear regression model is one of the primary techniques that is used for forecasting the stock market prices. In this regression model dataset acts as the prerequisite for price prediction. Based on the explanatory variables linear regression is mainly classified into two types ,such as simple linear regression and multiple linear regression. Linear regression model uses two different types such as Independent variables and Dependent variables named as x and y respectively(Wu et al. 2020). The equation is represented as $Y=\beta0+\beta1X$. $\beta0$ and $\beta1$ are the coefficients used for the variables. Artificial neural networks can be defined as one of the intelligent models in data mining techniques which can be used for predicting the stock market prediction.This Artificial neural networks includes a set of threshold functions. These functions are trained on the historical data and after connecting each other with adaptive weights, they are used to make future predictions(Bosco and Khan 2018).

The tool used to execute the program is Google colab and the databases are directly imported when the commands are instructed in the command prompt. Algorithm is implemented in the python code and the accuracy is obtained based on the dataset.

Linear Regression (LR) Algorithm

In Deep learning algorithms Linear Regression is based on the model of the relationship between two variables by fitting a linear equation to observe data, when one variable is considered to be an exploratory variable.

Pseudo Code

Input: Training Data
Output: Accuracy
Begin
    For all the existing data
    Read and Load the data set.
    Extract Data features of a sound file imported.

Add the technical indicators.
Initialize the LR classifier.
Train the classifier
Predict the test set
Return the Accuracy
End for

Artificial Neural Network (ANN) Algorithm:
Neural networks consist of a set of algorithms that certify the underlying relationships between the similar data. This algorithm is derived based on the human neural network.These include a set of threshold functions that are used for predicting the future stock prices.

Pseudo Code
Input: Training Data
Output: Accuracy
Begin
For all the existing data
Read and Load the data set.
Extract Data features of a sound file imported.
Add the technical indicators.
Initialize the ANN classifier.
Train the classifier
Predict the test set
Return the Accuracy
End for

Statistical Analysis
The analysis of stock price prediction was done using IBM SPSS software. Independent sample t test is carried out for analysis. Independent variables are dataset and dependent variable is accuracy(Ho 2017).

3. Results

The Artificial neural networks algorithm shows that databases are the abstract for determining the accurate predictions. Table 1 represents the prediction of the stock prices using Linear regression method.
In Table 2, The results achieved with p=1.000 (p>0.05) shows that two groups are statistically insignificant. Table 2 represents the group statistics analysis which include the total no.of. Values, mean, standard deviation, standard error mean. Linear regression method and SVM algorithm has an accuracy of 85% and 73% respectively. Standard error mean for ANN method is 1.035 less than Linear regression method.
Fig 1 represents the graph that explains the comparison of the accuracy value with the algorithm Linear regression method and Artificial neural networks algorithm, where the accuracy of Linear regression method is 85% and the accuracy value of the Artificial neural networks algorithm is 73%.

4. Discussion

The data evolution was performed using IBM SPSS version 21.To analyze the data, Independent sample T-test and group statistics can be carried out. In this research study using the novel approaches it is proved that the linear regression method has got the more accurate values than the Support vector Machine algorithm(Anand 2021).

Linear regression is a method which has lower time complexity when compared to other regression and classification methods(Bocklitz 2019). This method takes less time for compiling to get the accurate values(Shah, Campbell, and Zulkernine 2018). Linear regression performs exceptionally well for linearly separable data. This method is easier to implement, interpret and efficient to train. Whereas ANN model is a costly expensive model(Somers, Somers, and Bhattacharya 2016). ANN model works slow when it works on large datasets In ANN model the duration of network is unknown. The accuracy of output depends on the quality of the data(Damrongsakmethee and Neagoe 2020).

The limitations of the investors are the accurate stock price prediction and less efficiency. The main aim of the study is to provide accurate stock price prediction using linear regression method(Selvin et al. 2017). There are various applications and novel approaches where linear regression can be used such as predicting weather, forecasting results, academic reports(Selvin et al. 2017; Biswas et al. 2021).

5. Conclusion

In the proposed model using the novel approaches ,the accuracy percentage of predicting the stock market prices using linear regression model is 85%, whereas in the comparison model the Support vector machine algorithm has got the accuracy of 73% only.

Declarations
Conflict of Interests
No conflict of interest in this manuscript.

Authors Contribution
Author MKP was involved in data collection,data analysis, and manuscript writing. Author SAK was involved in the action process. Data verification and validation and Criteria 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.Oracle Tech Solutions Pvt.Ltd

2.Saveetha University

3.Saveetha Institute of Medical and Technical Sciences 4.Saveetha School of Engineering

6. References

Alfonso, Gerardo, A. Daniel Carnerero, Daniel R. Ramirez, and Teodoro Alamo. 2021. "Stock Forecasting Using Local Data." IEEE Access: Practical Innovations, Open Solutions 9: 9334–44.

Anand, C. 2021. "Comparison of Stock Price Prediction Models Using Pre-Trained Neural Networks." Journal of Ubiquitous Computing and Communication Technologies 3 (2): 122–34.

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.

Biswas, Milon, Atanu Shome, Md Ashraful Islam, Arafat Jahan Nova, and Shamim Ahmed. 2021. "Predicting Stock Market Price: A Logical Strategy Using Deep Learning." In 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). IEEE. https://doi.org/10.1109/iscaie51753.2021.9431817.

Bocklitz, Thomas. 2019. "Understanding of Non-Linear Parametric Regression and Classification Models: A Taylor Series Based Approach." Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods. https://doi.org/10.5220/0007682008740880.

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

Damrongsakmethee, Thitimanan, and Victor-Emil Neagoe. 2020. "Stock Market Prediction Using a Deep Learning Approach." In 2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). IEEE. https://doi.org/10.1109/ecai50035.2020.9223142.

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.

Erb, Claude B. n.d. "Has the Stock Market Been Overgrazed?" SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2432624.

Ho, Robert. 2017. "Hypothesis Testing: T Test for Independent and Correlated Groups." Understanding Statistics for the Social Sciences with IBM SPSS. https://doi.org/10.4324/9781315182452-13.

Jain, Sakshi, Neeraj Arya, and Shani Pratap Singh. 2020. "Stock Market Prediction Using Hybrid Approach." In Innovative Data Communication Technologies and Application, 476–88. Cham: Springer International Publishing.

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.

Matloff, Norman. 2017. "Linear Regression Models." Statistical Regression and Classification. https://doi.org/10.1201/9781315119588-2.

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.

Nayak, Aparna, M. M. Manohara Pai, and Radhika M. Pai. 2016. "Prediction Models for Indian Stock Market." Procedia Computer Science 89: 441–49.

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.

Poitras, Geoffrey. n.d. "Introduction: Stock Market Globalization, Past and Present." Handbook of Research on Stock Market Globalization. https://doi.org/10.4337/9780857938183.00008.

Selvin, Sreelekshmy, R. Vinayakumar, E. A. Gopalakrishnan, Vijay Krishna Menon, and K. P. Soman. 2017. "Stock Price Prediction Using LSTM, RNN and CNN-Sliding Window Model." In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE. https://doi.org/10.1109/icacci.2017.8126078.

Shah, Dev, Wesley Campbell, and Farhana H. Zulkernine. 2018. "A Comparative Study of LSTM and DNN for Stock Market Forecasting." In 2018 IEEE International Conference on Big Data (Big Data). IEEE. https://doi.org/10.1109/bigdata.2018.8622462.

Shaikh, Aryan, and Sandip Shinde. n.d. "Prediction of Stock Market Prices Using Prediction Algorithm and Sentiment Analysis." SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4111874.

Somers, P. A. A. M., P. A. A. Somers, and N. Bhattacharya. 2016. "A New Method for Processing Time Averaged Vibration Patterns: Linear Regression." Strain. https://doi.org/10.1111/str.12188.

Syed, Aamir Aijaz. 2021. "Symmetric and Asymmetric Influence of Macroeconomic Variables on Stock Prices Movement: Study of Indian Stock Market." New Challenges for Future Sustainability and Wellbeing. https://doi.org/10.1108/978-1-80043-968-920211017.

Upadhyay, Ved Prakash, Subhash Panwar, Ramchander Merugu, and Ravindra Panchariya. 2016. "Forecasting Stock Market Movements Using Various Kernel Functions in Support Vector Machine." In Proceedings of the International Conference on Advances in Information Communication Technology & Computing - AICTC '16. New York, New York, USA: ACM Press. https://doi.org/10.1145/2979779.2979886.

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.

Vohra, Archit A., and Paresh J. Tanna. 2021. "A Survey of Machine Learning Techniques Used on Indian Stock Market." IOP Conference Series. Materials Science and Engineering 1042 (1): 012021.

Wang, Yan, and Yuankai Guo. 2020. "Forecasting Method of Stock Market Volatility in Time Series Data Based on Mixed Model of ARIMA and XGBoost." China Communications 17 (3): 205–21.

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, Dingming, Xiaolong Wang, Jingyong Su, Buzhou Tang, and Shaocong Wu. 2020. "A Labeling Method for Financial Time Series Prediction Based on Trends." Entropy 22 (10). https://doi.org/10.3390/e22101162.

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. Group Statistics results (Mean of Linear Regression is 85 which is greater when compared to ANN Algorithm is 73 and Standard error mean for Linear Regression is 1.265 and ANN algorithm is 1.159)

	Algorithms	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Accuracy	Linear Regression	20	85	3.239	1.024
	ANN Algorithm	20	73	3.273	1.035

Table 2. The results achieved with p=1.000 (p>0.05) shows that two groups are statistically insignificant. The below table shows the results of independent variables of two algorithms with the comparison of accuracy and T-test for Equality of Mean.

		Levene's test for equality of variables		**T-test for Equality of Mean**						
		F	**Sig**	**t**	**df**	**Sig(2-tailed)**	**Mean difference**	**Std.Error difference**	**95% confidence interval of the difference**	
									Lower	**Upper**
Accuracy	**Equal variance assumed**	.000	1.000	8.379	18	0.000	12.200	1.456	9.141	15.259

	Equal variances not assumed			8.379	17.998	0.000	12.200	1.456	9.141	15.259

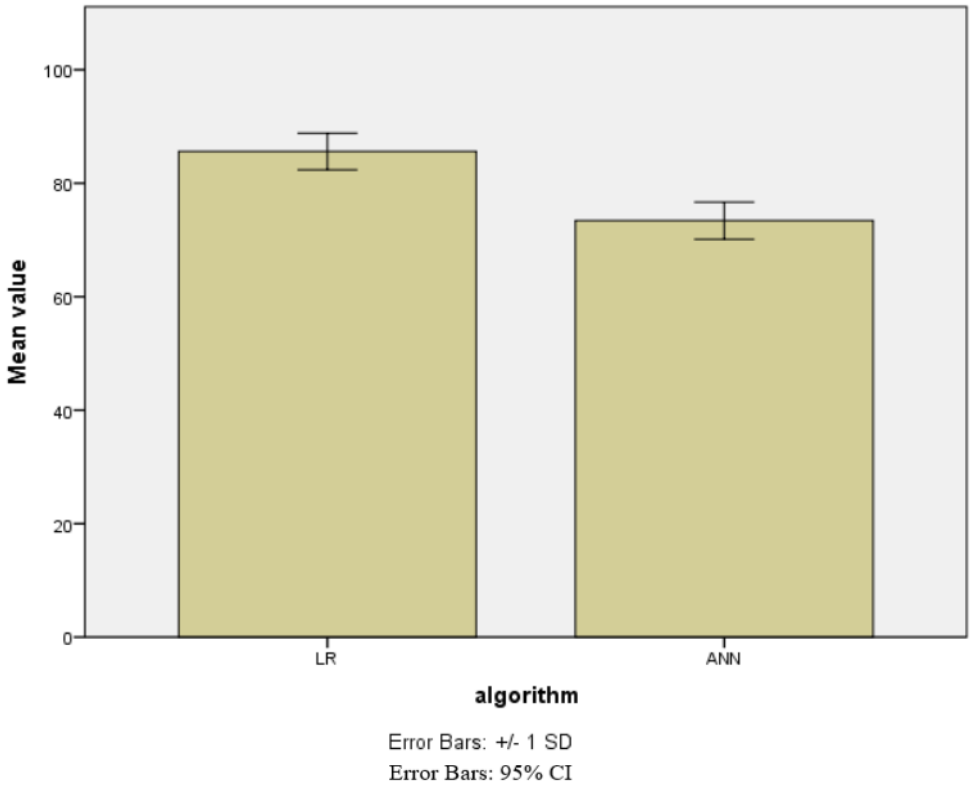

Fig. 1. Bar Chart representing the comparison of Mean Accuracy of Linear Regression and ANN Algorithm. Mean accuracy of Linear Regression is 85% appears to be better than ANN Algorithm which is 73%. The X-axis represents Linear Regression and ANN Algorithm and the Y-axis represents the mean accuracy ± 1 SD.