

Comparing neural network with linear regression for stock market prediction

Fachrul Kurniawan ^{a,1,*}, Yunifa Miftachul Arif ^{a,2}, Fresy Nugroho ^{a,3}, Mohammed Ikhlayel ^{b,4}

^a Department of Informatics Engineering, State Islamic University Maulana Malik Ibrahim Malang, Indonesia

^b Department of Information Technology and Communications, Al-Quds Open University, Palestine

¹ fachrulk@ti.uin-malang.ac.id; ² yunif4@ti.uin-malang.ac.id; ³ fresy@ti.uin-malang.ac.id; ⁴ miklil@qou.edu

* corresponding author

ARTICLE INFO

Article history

Received January 20, 2023

Revised February 3, 2023

Accepted February 24, 2023

Keywords

Neural network

Linear regression

Stock market

Prediction

ABSTRACT

There are both gains and losses possible in stock market investing. Brokerage firms' stock investments carry a higher risk of loss since their stock prices are not being tracked or analyzed, which might be problematic for businesses seeking investors or individuals. Thanks to progress in information and communication technologies, investors may now easily collect and analyze stock market data to determine whether to buy or sell. Implementing machine learning algorithms in data mining to obtain information close to the truth from the desired objective will make it easier for an individual or group of investors to make stock trades. In this study, we test hypotheses on the performance of a financial services firm's stock using various machine learning and regression techniques. The relative error for the neural network method is only 0.72 percentage points, while it is 0.78 percentage points for the Linear Regression. More training cycles must be applied to the Algoritma neural network to achieve more accurate results.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

Investors frequently choose stocks as their asset of choice because of the enormous potential rewards they provide [1]. On-the-go stock purchases and sales can be made through the use of a downloadable app. Seeds, StockBit, OctaFx, and a great number of additional applications, just like them, are all examples of stock programs for mobile devices [2], [3]. Investing in stocks comes with both positives and negatives that should be considered. Investing money in companies that do not have a track record of success or are accused of fraud can bring about financial devastation and bankruptcy [4], [5]. In addition to this, generating money on the stock market necessitates a high level of foresight in addition to an extensive understanding of the stock market.

Therefore, linear regression [6]–[8] and neural networks [9]–[11] are utilized in this study to produce projections regarding bank stocks. It is vital to develop simulations utilizing the tools that are accessible in order to optimize outcomes, especially buying and selling suggestions on the stock market [12]. This is done so investors can obtain a clear image through data trials using neural networks and linear regression techniques.

The data that was used in this investigation came from the Indonesian banking stake held by a single corporation. As a result of the size and reputation of the firm, this stock is now frequently recommended to investors for use as a bank service provider. This stock has a track record of providing superior returns to its shareholders.

2. Method

An early step in this study's contribution is developing a plan outlining the sequence of events that will occur throughout the experiment, allowing for more precise and repeatable results. This research follows the sequence shown below.

The research begins with defining the problem and developing a working hypothesis, followed by searching for relevant publications or studies. Second, it is time to gather some information. Information on a topic may be quickly found by conducting a web search. Yahoo Finance was utilized by the research team in this study for its extensive stock database.

In order to find an appropriate rhythm algorithm for use in forecasting, it is necessary first to identify the problem. This study chooses to employ linear regression analysis and a neural network. Both approaches can be helpful in data prediction and have their benefits. RapidMiner, a machine learning assistance program, is also utilized in this investigation.

Finally, the fourth step attempts to find answers to the issues identified. The answer is shown as a comparison of the accuracy of the linear regression technique and the neural network.

A neural network is a collection of interconnected computational nodes inspired by the structure of the human brain [13], [14]. ANN, like the brain's internal network, is made up of a collection of neurons that work together to process and convert incoming data [15], [16]. The term "weight" is used to describe this connection. The data collected is then sent into the propagation function's input and used to calculate the total load. As shown in Fig. 1, there are three stages in which data will travel: the input, the hidden layer, and the output.

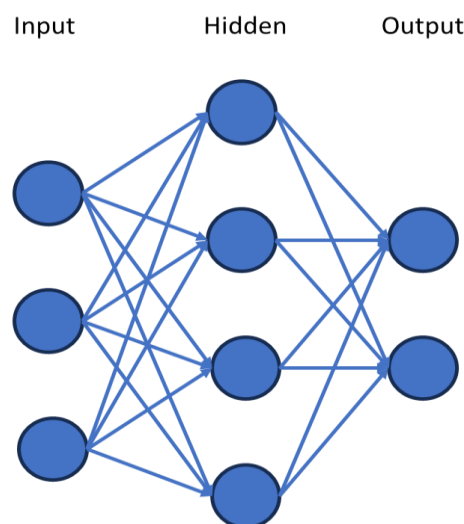


Fig. 1. Neural network

Linear Regression is divided into two parts: simple and multiple linear regression. Simple linear regression is an equation model that uses the relationship of one independent variable/predictor (X) with a dependent variable/response (Y) [15]. The difference in the multiple linear regression method is that the independent variables have more than one variable [16]. The equation used in linear Regression is as follows. $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots$

3. Results and Discussion

This project makes use of existing resources by first developing a rapidminer software process schematic, as seen in Fig. 2. Rapidminer provides a simple step by running the process scheme that has been made to see the prediction results using the neural network method. This neural network still uses training cycles = 200 and learning rate = 0.01, which is the default setting from the rapidminer. Three tests will be proposed with different training cycles and learning rates to improve accuracy. Fig. 3 show the neural network architecture for testing process. Fig. 4 shows an example of the prediction results performed by a neural network.

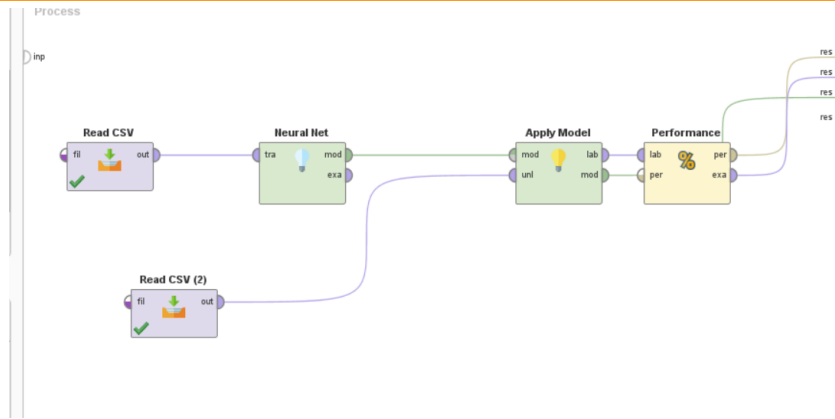


Fig. 2. Neural network process

The input is selected on the neurons, namely low, which goes to nodes in hidden layers 2 and 4. In this test, it can be seen that linear Regression still has higher accuracy than the neural network. However, the accuracy of neural networks can still be improved by changing the training cycles and learning rates used. Therefore, the second test scenario is carried out by changing the training cycles to 500, and the learning rate to 0.1.

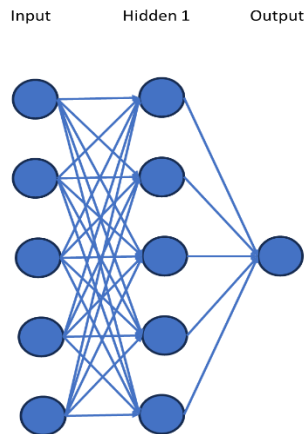


Fig. 3. Neural network architecture for testing

Row No.	Close	prediction(C...	Date	Open	High	Low
1	7600	7567.865	Jan 10, 2022	7575	7650	7550
2	7700	7619.106	Jan 11, 2022	7650	7700	7625
3	7700	7620.166	Jan 12, 2022	7700	7700	7650
4	7700	7643.822	Jan 13, 2022	7700	7750	7650
5	7850	7690.531	Jan 14, 2022	7700	7850	7650
6	7750	7684.359	Jan 17, 2022	7900	7900	7700
7	7675	7563.066	Jan 18, 2022	7775	7775	7550
8	7675	7597.293	Jan 19, 2022	7650	7725	7575
9	7775	7631.137	Jan 20, 2022	7725	7775	7625
10	7950	7746.674	Jan 21, 2022	7775	7950	7700
11	7800	7735.174	Jan 24, 2022	7825	7850	7775
12	7775	7647.535	Jan 25, 2022	7675	7775	7625
13	7700	7644.827	Jan 26, 2022	7700	7750	7650
14	7800	7675.280	Jan 27, 2022	7625	7800	7625
15	7775	7681.988	Jan 28, 2022	7800	7825	7700
16	7625	7615.833	Jan 31, 2022	7775	7775	7625
17	7800	7661.813	Feb 2, 2022	7775	7800	7675
18	7725	7682.455	Feb 3, 2022	7800	7825	7700

Fig. 4. Prediction results of neural network test 1

Open, high, and low inputs are linked to hidden 1 in neural network test 2. In addition, the output is linked to nodes 1, 2, 3, and 4. All three types of inputs—open, high, and low—are sent into the hidden layer's nodes.

Table 1 summarizes the error performance outcomes from each approach, allowing us to compare their relative precision. The table shows that the third neural network test also has less error than the linear Regression. Overall, the bigger the number of training cycles, the better the results. As a result, the margin of error is decreasing, and precision is improving.

Table.1 Error performance

Method	RMSE	Absolute error	Relative error
Regresi Linear	80.380 +/- 0.000	61.263 +/- 52.036	0.78% +/- 0.66%
Neural network (test 1)	103.962 +/- 0.000	88.848 +/- 53.982	1.13% +/- 0.67%
Neural network (test 2)	91.101 +/- 0.000	76.236 +/- 49.876	0.97% +/- 0.62%
Neural network (test 3)	67.042 +/- 0.000	56.100 +/- 36.707	0.72% +/- 0.47%

Linear regression processing is seen in Fig. 5. Find RMSE, absolute, and relative errors with apply model and performance first. Fig. 6 displays the results of the low attribute prediction of stocks, which has a coefficient of 0.977, an error of 0.005, and a standard coefficient of 0.994.

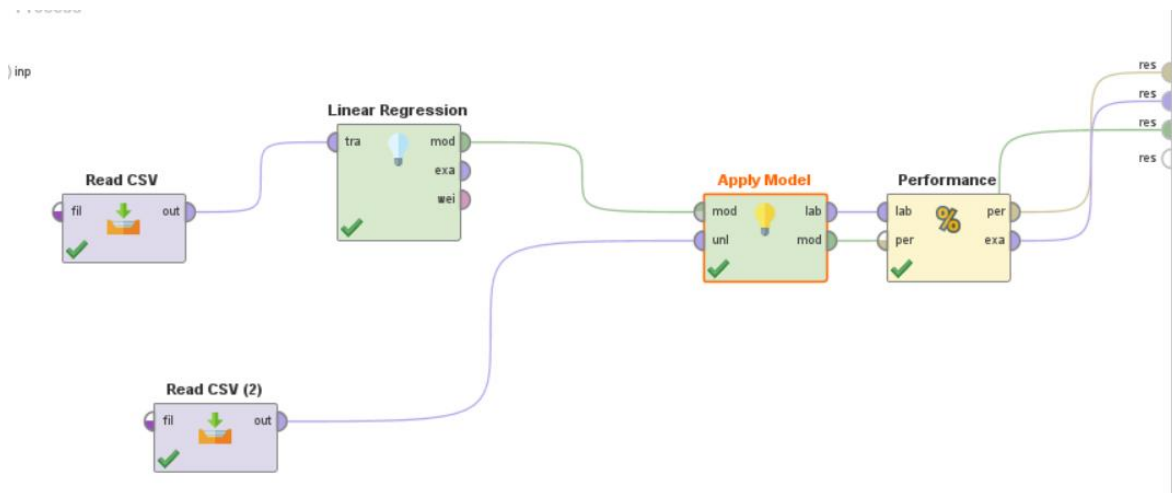


Fig. 5. Linear regression process

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Low	0.977	0.005	0.994	1	194.416	0	****
(Intercept)	211.955	31.881	?	?	6.648	0.000	****

Fig. 6. Linear regression prediction results

Linear Regression is a straightforward method that may be quickly applied to produce reliable outcomes. In contrast to other complicated algorithms, training these models requires relatively little computing power. Therefore, they may be used even on systems with limited resources. Linear Regression's temporal complexity is far lower than other machine learning techniques. Linear Regression's mathematical formulae are also easy to grasp and apply. Therefore, learning linear Regression is a breeze.

Finding the nature of the association between variables is a common use of linear Regression because of its near-perfect fit to linearly separable datasets. Overfitting can also be mitigated by regularization. When a machine learning model catches the noisy data along with the clean, it is said to have overfit the dataset, lowering the model's quality and its results on the test data [17]. A regularization is a straightforward approach that can effectively reduce a function's complexity to mitigate overfitting.

Linear Regression is problematic for complicated datasets because it presumes a linear connection between input and output variables. In most situations, a straight line does not provide the best fit for the data since the connection between the variables in the dataset is not linear [18]. In some cases, a more sophisticated function can more accurately capture the information, which is why linear regression models typically have poor predictive power [19]. It is essential to adequately handle outliers before applying linear Regression to a dataset since they might significantly affect the results.

On the other hand, neural networks initially improve visual analysis. Artificial neural networks can accomplish more complicated tasks than other machines because they are akin to human brain networks. Neural networks can analyze disorganized data, which is another benefit [20]. ANNs make organizing unstructured data simpler.

Neural networks' adaptability is its third benefit. For any purpose, an ANN changes its structure. Neural networks may transition from machine learning to complicated applications. Unlike many machine learning techniques and applications, this is flexible. Artificial neural networks swiftly adapt to new surroundings and show their talents, suggesting that these networks require less and more flexible training. Computation requires no involvement.

4. Conclusion

According to the findings of this study, linear Regression and neural networks may be able to be used to make accurate predictions on the stock market. When compared to the relative error number produced by linear Regression, the neural network technique produces just 0.72% of it, whereas linear Regression produces 0.78%. Additional training cycles are required in order to increase the accuracy of the predictions made by the neural network. In the not-too-distant future, researchers hope to compare the precision of this method with that of many other machine-learning strategies. The combination of clustering, classification, and association might be interesting for future implementations.

References

- [1] J. Y. Campbell and L. M. Viceira, "Strategic Asset Allocation." *Oxford University Press Oxford*, pp. 381-420, 2002, doi: [10.1093/0198296940.001.0001](https://doi.org/10.1093/0198296940.001.0001).
- [2] M. Yusuf, M. Rahmani, A. Fakultaspascasarjana, and U. A. Banjarmasin, "Sharia Law Analysis of Binary Option," *Syariah J. Huk. dan Pemikir.*, vol. 22, no. 2, pp. 141-149, Dec. 2022, Accessed: Dec. 21, 2022. [Online]. Available: <https://jurnal.uin-antasari.ac.id/index.php/syariah/article/view/6454>.
- [3] R. Chaysiri and C. Ngauv, "Prediction of Closing Stock Prices Using the Artificial Neural Network in the Market for Alternative Investment (MAI) of the Stock Exchange of Thailand (SET)," in *Integrated Uncertainty in Knowledge Modelling and Decision Making*, Thammasat University, 2020, pp. 335-345, doi: [10.1007/978-3-030-62509-2_28](https://doi.org/10.1007/978-3-030-62509-2_28).
- [4] A. R. Admati, "A Skeptical View of Financialized Corporate Governance," *J. Econ. Perspect.*, vol. 31, no. 3, pp. 131-150, Aug. 2017, doi: [10.1257/jep.31.3.131](https://doi.org/10.1257/jep.31.3.131).
- [5] H. Grove and M. Clouse, "Financial and Non-Financial Fraud Risk Assessment," *J. Forensic Investig. Account.*, vol. 12, no. 3, p. 2020, 2020. [Online]. Available at: <http://web.nacva.com.s3.amazonaws.com/JFIA/Issues/JFIA-2020-No3-3.pdf>.
- [6] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 1, no. 4, pp. 140-147, 2020, doi: [10.38094/jastt1457](https://doi.org/10.38094/jastt1457).
- [7] A. Sharma, D. Bhuriya, and U. Singh, "Survey of stock market prediction using machine learning approach," in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, Apr. 2017, pp. 506-509, doi: [10.1109/ICECA.2017.8212715](https://doi.org/10.1109/ICECA.2017.8212715).
- [8] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices," *J. Financ. Data Sci.*, vol. 4, no. 3, pp. 183-201, Sep. 2018, doi: [10.1016/j.jfds.2018.04.003](https://doi.org/10.1016/j.jfds.2018.04.003).
- [9] Ö. Ican and T. B. Çelik, "Stock Market Prediction Performance of Neural Networks: A Literature Review," *Int. J. Econ. Financ.*, vol. 9, no. 11, p. 100, Oct. 2017, doi: [10.5539/ijef.v9n11p100](https://doi.org/10.5539/ijef.v9n11p100).

-
- [10] M. F. Masouleh and A. Bagheri, "Forecasting Stock Exchange Data using Group Method of Data Handling Neural Network Approach," *Knowl. Eng. Data Sci.*, vol. 4, no. 1, p. 1, Aug. 2021, doi: [10.17977/um018v4i12021p1-13](https://doi.org/10.17977/um018v4i12021p1-13).
- [11] Y.-G. Song, Y.-L. Zhou, and R.-J. Han, "Neural networks for stock price prediction," *arXiv Prepr. arXiv1805.11317*, pp. 1-13, 2018. [Online]. Available at: <https://arxiv.org/abs/1805.11317>.
- [12] S. Sohngir, D. Wang, A. Pomeranets, and T. M. Khoshgoftaar, "Big Data: Deep Learning for financial sentiment analysis," *J. Big Data*, vol. 5, no. 1, p. 3, Dec. 2018, doi: [10.1186/s40537-017-0111-6](https://doi.org/10.1186/s40537-017-0111-6).
- [13] A. Azhari, A. Susanto, A. Pranolo, and Y. Mao, "Neural Network Classification of Brainwave Alpha Signals in Cognitive Activities," *Knowl. Eng. Data Sci.*, vol. 2, no. 2, p. 47, 2019, doi: [10.17977/um018v2i22019p47-57](https://doi.org/10.17977/um018v2i22019p47-57).
- [14] E. Vocaturro and P. Veltri, "On the use of Networks in Biomedicine," *Procedia Comput. Sci.*, vol. 110, pp. 498–503, Jan. 2017, doi: [10.1016/j.procs.2017.06.132](https://doi.org/10.1016/j.procs.2017.06.132).
- [15] A. Pyataeva and A. Dzyuba, "Artificial neural network technology for lips reading," *E3S Web Conf.*, vol. 333, p. 01009, Dec. 2021, doi: [10.1051/e3sconf/202133301009](https://doi.org/10.1051/e3sconf/202133301009).
- [16] A. P. Wibawa, W. Lestari, A. B. P. Utama, I. T. Saputra, and Z. N. Izdihar, "Multilayer Perceptron untuk Prediksi Sessions pada Sebuah Website Journal Elektronik," *Indones. J. Data Sci.*, vol. 1, no. 3, Dec. 2020, doi: [10.33096/ijodas.v1i3.15](https://doi.org/10.33096/ijodas.v1i3.15).
- [17] C. Renggli, L. Rimanic, N. M. Gürel, B. Karlaš, W. Wu, and C. Zhang, "A data quality-driven view of mlops," *arXiv Prepr. arXiv2102.07750*, pp. 1-12, 2021. [Online]. Available at: <https://arxiv.org/abs/2102.07750>.
- [18] J. Gauthier, Q. V. Wu, and T. A. Gooley, "Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians," *Bone Marrow Transplant.*, vol. 55, no. 4, pp. 675–680, Apr. 2020, doi: [10.1038/s41409-019-0679-x](https://doi.org/10.1038/s41409-019-0679-x).
- [19] V. Stanev *et al.*, "Machine learning modeling of superconducting critical temperature," *npj Comput. Mater.*, vol. 4, no. 1, p. 29, Jun. 2018, doi: [10.1038/s41524-018-0085-8](https://doi.org/10.1038/s41524-018-0085-8).
- [20] J. Santoso, E. I. Setiawan, F. X. Ferdinandus, G. Gunawan, and L. Hernandez, "Indonesian Language Term Extraction using Multi-Task Neural Network," *Knowl. Eng. Data Sci.*, vol. 5, no. 2, p. 160, Dec. 2022, doi: [10.17977/um018v5i22022p160-167](https://doi.org/10.17977/um018v5i22022p160-167).
-