Scientific Journal of Informatics
Vol. 11, No. 1, Feb 2024
p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

Optimization Selection on Deep Learning Algorithm for Stock Price Prediction in Indonesia Companies

Gunawan[1*], Wresti Andriani[2], Sawaviyya Anandianskha[3], Aang Alim Murtopo[4], Bangkit Indarmawan Nugroho[5], Naella Nabila Putri Wahyuning Naja[6]
[1,2,3,4]Department of Informatic Engineering, STMIK YMI Tegal, Indonesia
[5]Department of Information System, STMIK YMI Tegal, Indonesia
[6]Department of Management, Universitas Negeri Semarang, Indonesia

Abstract.
Purpose: Share price movements after the COVID-19 pandemic experienced a decline in several sectors, especially in the share prices of the Aneka Tambang Company, which operates in the mining sector, the Wijaya Karya Company in the construction sector, and the Sinar Mas Company, which is a Holding Company. Several factors influence this, including investors' hesitation in investing their money. This research aims to predict stock price movements using a Deep Learning algorithm, which is optimized using Selection optimization at three large companies in Indonesia, namely PT. ANTAM, PT. WIKA, and PT. SINAR MAS. So that it can provide the correct information to investors to avoid losses.
Method: research through collecting data from the three companies, preprocessing, and then analyzing research data with several alternatives. The combination of inputs from the three companies using the deep learning method is then optimized using selection optimization to produce the best accuracy and use the results of the RMSE evaluation.
Results: The results of this research show that by using the Deep Learning method, the best evaluation results were obtained for the Company PT Wijaya Karya with an RMSE value of 0.432, an MAE value of 0.31505 and an MSE value of 1913.953. These results were then optimized using Selection optimization to obtain an RMSE increase of 0.022, namely 0.410.
Novelty: The contribution of this research is to get the best combination of input variables obtained using the windowing process from the three companies, which are then processed using the Deep Learning method to produce the most accurate evaluation results from the three companies, then the results are optimized again using Selection optimization to get the more optimal results.

Keywords: Deep learning, Input variable selection, Selection optimization, Stock price
Received October 2023 / **Revised** November 2023 / **Accepted** February 2024

INTRODUCTION

Stock investment activities in the post-pandemic period are things that make investors hesitate because there are still many companies that are still recovering from the downturn [1]. Stock price movements fluctuate and are difficult to predict, making it challenging to predict ups and downs [2]. Several factors can make investors do activities on stocks, and some even make it as income [3]. These factors include external and internal factors; some even connect with weather factors and investor moods in increasing or decreasing stock prices [4].

These internal variables include operational expenses, while external factors include net interest margin, return on equity, and return on assets [5]. In manufacturing companies, internal factors are influenced by micro factors, namely marketing announcements, financing, management announcements of directors, diversification announcements, investment announcements, employment announcements, and financial announcements of the Company, which are generally affected due to the COVID-19 pandemic [6], [7]. At the same time, external factors are influenced by external factors [8], such as announcements of a country's government, legal announcements, industry announcements, the political turmoil of a country, fluctuations in currency values, and various issues from outside and within the country [9].

*Corresponding author.
Email addresses: gunawan.gayo@gmail.com (Gunawan)
DOI: 10.15294/sji.v11i1.47935

The movement of a company's stock price powerfully provokes investors to examine carefully and carefully so that the decision to buy and sell a form of security can generate profits and avoid losses, even though internal factors also affect stock price movements [10]. This is one of the reasons why share price movements cannot be predicted with certainty. So, it is essential to research to minimize losses.

Deep learning is a computational technology to change paradigms in data processing, pattern recognition, and understanding complex systems [11], [12], [13], [14]. This method is based on the architecture of deep artificial neural networks [15], inspired by the structure and function of human neural networks [16]. These artificial neural networks consist of many layers (called hidden layers") [17], [18], which allows these algorithms to extract increasingly complex features from input data. This significantly contrasts conventional methods, where feature extraction is generally done manually.

Several studies predict the movement of a company's stock price [19], predictions are made using SVM and generate RMSE of 20,281, as well as research [20] which examines stock prices of shipping companies using linear regression and produces RMSE of 7,522, as well as research [1], [21], [22], [23] which predicts the sentiment of analysts on stock prices that use SVM which produces low data.

In the previous research above, there are weaknesses, for example, in the linear regression method to prediction, which is that this method is only limited to linear relationships but cannot show non-linear relationships without additional transformations that combine non-linear terms [24], [25]. At the same time, the Support Vector Machine has the disadvantage that it is unsuitable for large data sets because it requires a long training time [26], [27], [28], [29]. Some studies have proven that the evaluation results using the RMSE Support Vector Machine are still more significant than those using Linear Expression [30], [31].

This study will use Deep Learning algorithms that are optimized using Selection optimization to predict stock price movements in three major companies in Indonesia, namely PT. ANTAM [32], PT. WIKA [33], and PT. SINAR MAS [34] will be affected by the COVID-19 pandemic. This study aims to help investors predict stock price movements to avoid losses and determine whether this algorithm can predict stock price movements well, produce good accuracy, and be more optimal after being optimized using Selection Optimization.

METHODS

The research method that was carried out appears in Figure 1.

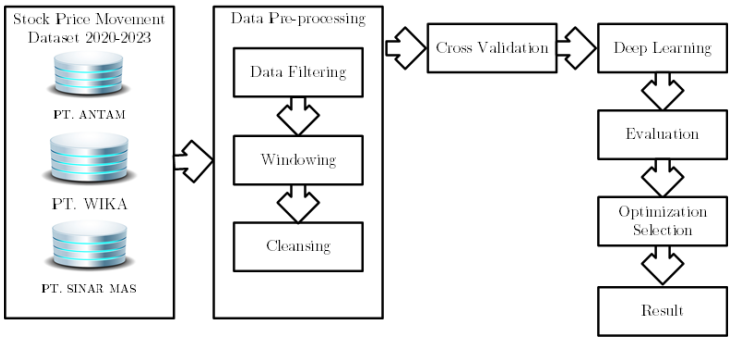


Figure 1. Research methods

Figure 1 shows that the first step of this study is data collection or data collection. This data is obtained online from the Yahoo.com page of stock price movements of three major companies in Indonesia. PT. ANTAM, PT. WIKA, and PT. SINAS MAS From Historical data from 2020 to [35]–[37]. The results obtained are in Excel form. This period was taken because, after the coronavirus pandemic in 2020, there must have been many companies that were affected and until now are still not healthy, bounce back. The data generated was 73 from each Company.

Data preprocessing is done after getting the dataset used [38], including filtering by changing the data arrangement from 2023 to 2020. The data variables obtained are seven: open, high, low, close, adj close,

and volume. Then, define the target or data label. This study focused on the Close variable. Close Price is the closing price of a stock [39], which measures several pieces of information that affect investor decisions at the close of trading and the Price of the following stock [40].

The next activity is windowing data [41], which converts univariate data from target data into multivariate data. The technique is sliding windows with daily or daily windowing size [42].

The next step is dividing training data and testing data [43]. This division is done with the 10-fold cross-validation model, a nested operator [44]. The validation concept divides the training data into ten equal parts and learns ten times [45]. Each time, it is selected on another part of the data set for learning, as many as nine, and the rest are used as testing. Then, the average and deviation values of 10 test results were calculated [46].

The next stage is to choose an evaluation model to measure the accuracy of the deep learning algorithm using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), where the lower the value, the better the level of accuracy. The accuracy of predictions is determined from the smallest value of each data accuracy method [47].

RESULTS AND DISCUSSIONS

The data to be used as a data set in this study is taken online from Yahoo.com on the stock data of three companies in Indonesia, namely PT. ANTAM, PT. WIKA, and PT. SINAR MAS, in Historical Data from 2020 to 2023, as many as 730. Data are obtained in Table 2. This research uses a computer with a Core-i7 processor specification, 512 GB SSD, 16 GB memory, and a 21-inch monitor. The software used is RapidMiner Studio version 10.

Table 1. Stock price of companies in Indonesia

Date	PT. ANTAM						PT. WIKA						PT. SINAR MAS					
	Open	High	Low	Close	Adj Close	Vol.	Open	High	Low	Close	Adj Close	Vol.	Open	High	Low	Close	Adj Close	Vol.
8/30/23	412	418	402	404	404	16526600	2010	2010	1995	1995	1995	31964100	4410	4420	4380	4400	4400	26300
8/29/23	378	414	374	412	412	47442600	1995	2000	1980	1995	1995	23010600	4380	4450	4380	4410	4410	7100
8/28/23	378	380	372	374	374	4604200	1995	2010	1980	1995	1995	20521600	4410	4460	4380	4380	4380	19200
8/25/23	388	388	376	378	378	3892100	1995	2010	1980	1990	1990	34868200	4430	4430	4400	4400	4400	4700
8/24/23	384	384	376	378	378	5639000	1975	2020	1975	1995	1995	81402400	4440	4490	4430	4430	4430	23700
8/23/23	376	382	376	382	382	6885500	1950	1965	1950	1955	1955	35653600	4480	4480	4400	4400	4400	30500
8/22/23	376	378	370	376	376	5894900	1950	1960	1935	1945	1945	36188400	4400	4450	4400	4450	4450	7300
8/21/23	384	386	372	372	372	9363100	1950	1965	1950	1950	1950	18371900	4410	4420	4400	4400	4400	14400
8/18/23	384	388	380	384	384	4112900	1960	1970	1945	1950	1950	42360900	4400	4420	4370	4410	4410	8400
8/16/23	392	392	380	384	384	13612900	1985	1990	1960	1960	1960	62356300	4400	4440	4390	4400	4400	6700
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8/31/20	1300	1305	1220	1240	1240	28789500	830	855	810	820	769.7685	229956400	3200	3260	3180	3180	2571.635	800

The dataset in Table 1 is obtained from the scribbling results on the page PT. ANTAM, PT. WIKA, and PT. SINAS MAS [35]–[37]. The data is then prepared to become data ready to be processed through data preprocessing. The variables obtained consist of 7 data: Open, High, Low, Close, Adj Close, and Volume. This study used close variables as targets or labels, as shown in Table 2.

Table 2. Variable close of all three companies

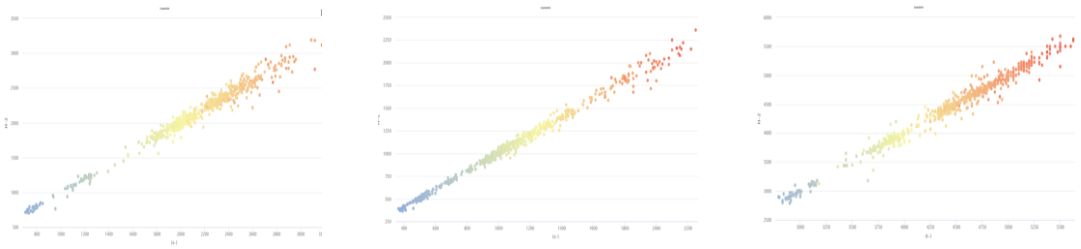
Date	Close		
	PT. ANTAM	PT. WIKA	PT. SINAR MAS
8/30/23	1995	404	4400
8/29/23	1995	412	4410
8/28/23	1995	374	4380
8/25/23	1990	378	4400
8/24/23	1995	378	4430
8/23/23	1955	382	4400
8/22/23	1945	376	4450
8/21/23	1950	372	4400
8/18/23	1950	384	4410
8/16/23	1960	384	4400
⋮	⋮	⋮	⋮
8/31/20	820	1240	3180

By using the close variable as the target, a sliding window process is carried out, namely to convert univariate data into multivariate data by making data before the target data into event data the day before (H-1), two days back (H-2) and three days earlier (H-3), into data that will become training data, the results are as in Table 3.

Table 3. Sliding window results

	PT. ANTAM			PT. WIKA				PT. SINAR MAS			
Close	H-1	H-2	H-3	Close	H-1	H-2	H-3	Close	H-1	H-2	H-3
1995	1995	1995	1990	404	412	374	378	4400	4410	4380	4400
1995	1995	1990	1995	412	374	378	378	4410	4380	4400	4430
1995	1990	1995	1955	374	378	378	382	4380	4400	4430	4400
1990	1995	1955	1945	378	378	382	376	4400	4430	4400	4450
1995	1955	1945	1950	378	382	376	372	4430	4400	4450	4400
1955	1945	1950	1950	382	376	372	384	4400	4450	4400	4410
1945	1950	1950	1960	376	372	384	384	4450	4400	4410	4400
1950	1950	1960	1990	372	384	384	392	4400	4410	4400	4390
1950	1960	1990	2000	384	384	392	384	4410	4400	4390	4400
1960	1990	2000	1990	384	392	384	388	4400	4390	4400	4420
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
830	850	840	820	1225	1250	1250	1240	3110	3160	3180	3180

After obtaining sliding window result data such as Table 3, the data is ready to be processed using Deep learning methods to find the most optimal comparison of accuracy results MAE and RMSE (Root Mean Square Error) (Mean Absolute Error). Graphic Image of PT. ANTAM, PT. WIKA and PT. SINAR MAS as shown in Figure 2.



(a) (b) (c)

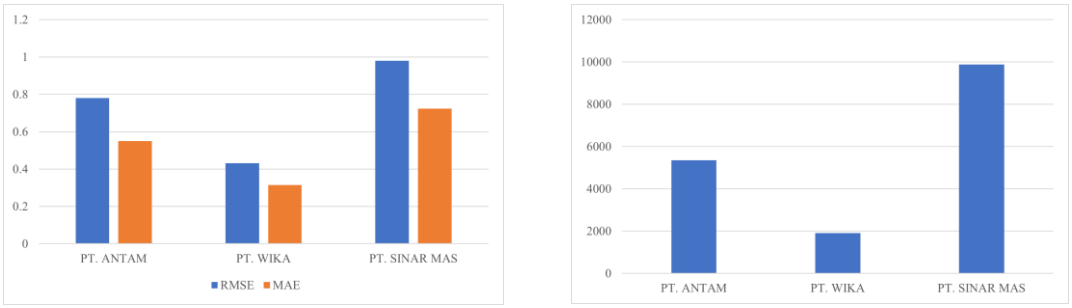
Figure 2. Stock price movements of (a) PT. ANTAM, (b) PT. WIKA and (c) PT. SINAR MAS

Table 4 displays the findings of evaluating the three businesses' deep learning techniques.

Table 4. Comparison of deep learning methods

Company	RMSE	MAE	MSE
PT. ANTAM	0.78973	0.54948	6355.132
PT. WIKA	0.43235	0.31505	1913.953
PT. SINAR MAS	0.98329	0.72337	9885.218

From Table 4, the RMSE evaluation results at PT can be seen. WIKA has the best accuracy, 0.432235, MAE of 0.31505, and MSE of 1913.953, followed by PT. ANTAM with RMSE of 0.78973, MAE of 0.54948, MSE of 6355.132, and PT. SINAR MAS, RMSE of 0.98329, MAE of 0.72337, and MSE of 9885.218. The results of the Table Graph in Deep Learning Table 4 are shown in Figure 3.



(a) (b)

Figure 3. Profound learning comparison results (a) RMSE and MAE, (b) MSE

Then, the results of PT. WIKA's RMSE evaluation was optimized using selection optimization, and the evaluation results were increased to Table 5.

Table 5. Optimization selection PT. WIKA results

Item	Results
RMSE	0.41016
MAE	0.27990
MSE	1741.10200

From the research that has been carried out, it can be seen that using the Deep Learning method, the results obtained were the share price movements of PT. WIKA received the best evaluation results: RMSE of 0.43235, MAE of 0.31505, and MSE of 1913.95. After optimization using Optimization Selection, even better evaluation results were obtained. Namely, RMSE of 0.41016, MAE of 0.27990, and MSE of 1741.10, the difference obtained from the RMSE value is 0.022, MAE is 0.03515, and MSE is 172.85, this indicates that the share price movement of PT. WIKA is the safest to invest. So that investors can safely invest and avoid the risk of loss.

CONCLUSION

Based on the research that has been carried out, it can be concluded that the share price movements of the three companies in Indonesia, the share prices of PT. WIKA is the best and safest investment because the evaluation results are better than PT's. SINAR MAS and PT. ANTAM has an RMSE of 0.4325, MAE of 0.315, and MSE of 1913.95. Optimized Optimization Selection produces an even better evaluation value with a value difference of 0.02, an RMSE value of 0.41016, an MAE value of 0.27990, and an MSE value of 1741.10. So, it can be used as information for investors to prefer PT. WIKA. Regarding the level of security in investing, this research aims to determine stock price movements after the coronavirus pandemic. The research results show that the Deep Learning approach has the best accuracy at PT. WiKA 0.02. To make it easier for investors planning to invest in this third business to decide whether to buy or sell PT. WIKA shares. Furthermore, this research only carries out experiments using Deep Learning algorithms, which are optimized using selection optimization, so further research can be carried out on applying other feature selection optimization algorithms to test whether this Deep Learning algorithm can increase its accuracy value.

REFERENCES

[1] B. Sulistio and D. Suhartono, "Utilizing Bert and Cnn-Lstm in Stock Price Prediction Using Data Sentiment Analysis and Technical Analysis of Stock and Commodity," *ICIC Express Lett.*, vol. 17, no. 2, pp. 171–179, 2023, doi: 10.24507/icicel.17.02.171.

[2] F. Ronaghi, M. Salimibeni, F. Naderkhani, and A. Mohammadi, "COVID19-HPSMP: COVID-19 adopted Hybrid and Parallel deep information fusion framework for stock price movement prediction," *Expert Syst. Appl.*, vol. 187, p. 115879, 2022, doi: https://doi.org/10.1016/j.eswa.2021.115879.

[3] Q. Wu, X. Liu, J. Qin, L. Zhou, A. Mardani, and M. Deveci, "An integrated generalized TODIM

model for portfolio selection based on financial performance of firms," *Knowledge-Based Syst.*, vol. 249, 2022, doi: 10.1016/j.knosys.2022.108794.
[4] W. Tarczyński, U. Mentel, G. Mentel, and U. Shahzad, "The influence of investors' mood on the stock prices: Evidence from energy firms in warsaw stock exchange, poland," *Energies*, vol. 14, no. 21, 2021, doi: 10.3390/en14217396.
[5] I. E. Ahmed, R. Mehdi, and E. A. Mohamed, "The role of artificial intelligence in developing a banking risk index: an application of Adaptive Neural Network-Based Fuzzy Inference System (ANFIS)," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13873–13895, 2023, doi: 10.1007/s10462-023-10473-9.
[6] I. M. García-Sánchez and A. García-Sánchez, "Corporate social responsibility during COVID-19 pandemic," *J. Open Innov. Technol. Mark. Complex.*, vol. 6, no. 4, pp. 1–21, 2020, doi: 10.3390/joitmc6040126.
[7] L. P. Ling and Y. Dasril, "Portfolio Selection Strategies in Bursa Malaysia Based on Quadratic Programming," *J. Inf. Syst. Explor. Res.*, vol. 1, no. 2, pp. 93–102, 2023, doi: 10.52465/joiser.v1i2.178.
[8] N. Linder, S. Rosenthal, P. Sörqvist, and S. Barthel, "Internal and External Factors' Influence on Recycling: Insights From a Laboratory Experiment With Observed Behavior," *Front. Psychol.*, vol. 12, no. July, pp. 1–12, 2021, doi: 10.3389/fpsyg.2021.699410.
[9] B. Eichengreen, R. Hausmann, and U. Panizza, *Yet it Endures: The Persistence of Original Sin*, vol. 34, no. 1. Springer US, 2023. doi: 10.1007/s11079-022-09704-3.
[10] V. K. Sen Liew, "The effect of novel coronavirus pandemic on tourism share prices," *J. Tour. Futur.*, vol. 8, no. 1, pp. 109–124, 2022, doi: 10.1108/JTF-03-2020-0045.
[11] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-Learning in Neural Networks: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5149–5169, 2022, doi: 10.1109/TPAMI.2021.3079209.
[12] X. Liu *et al.*, "Self-Supervised Learning: Generative or Contrastive," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 857–876, 2023, doi: 10.1109/TKDE.2021.3090866.
[13] M. Endah H, R. N. Wijaya, and H. Khotibul Ahsan, "Enhancing cirrhosis detection: A deep learning approach with convolutional neural networks," *J. Soft Comput. Explor.*, vol. 4, no. 4, pp. 195–203, 2023, doi: 10.52465/joscex.v4i4.226.
[14] D. H. Fudholi, S. Rani, D. M. Arifin, and M. R. Satyatama, "Deep Learning-based Mobile Tourism Recommender System," *Sci. J. Informatics*, vol. 8, no. 1, pp. 111–118, 2021, doi: 10.15294/sji.v8i1.29262.
[15] S. Kim *et al.*, "Integration of neural network-based symbolic regression in deep learning for scientific discovery," *IEEE Trans. neural networks Learn. Syst.*, vol. 32, no. 9, pp. 4166–4177, 2020.
[16] A. Mehonic, A. Sebastian, B. Rajendran, O. Simeone, E. Vasilaki, and A. J. Kenyon, "Memristors—From in-memory computing, deep learning acceleration, and spiking neural networks to the future of neuromorphic and bio-inspired computing," *Adv. Intell. Syst.*, vol. 2, no. 11, p. 2000085, 2020.
[17] X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," *Int. J. Mach. Learn. Cybern.*, vol. 11, pp. 747–750, 2020.
[18] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Syst. Appl.*, vol. 166, p. 114060, 2021, doi: 10.1016/j.eswa.2020.114060.
[19] H. Jin, L. Wang, Z. Xiao, and H.-G. Fung, "What firm risk factors drive bank loan pricing and other terms? Evidence from China," *Account. Financ.*, vol. 63, no. 3, pp. 2985–3010, Sep. 2023, doi: https://doi.org/10.1111/acfi.13001.
[20] A. Rai, S. R. Luwang, M. Nurujjaman, C. Hens, P. Kuila, and K. Debnath, "Detection and forecasting of extreme events in stock price triggered by fundamental, technical, and external factors," *Chaos, Solitons & Fractals*, vol. 173, p. 113716, 2023, doi: https://doi.org/10.1016/j.chaos.2023.113716.
[21] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," *Proc. - 5th IEEE Int. Conf. Big Data Serv. Appl. BigDataService 2019, Work. Big Data Water Resour. Environ. Hydraul. Eng. Work. Medical, Heal. Using Big Data Technol.*, pp. 205–208, 2019, doi: 10.1109/BigDataService.2019.00035.
[22] T. Sidogi, R. Mbuvha, and T. Marwala, "Stock price prediction using sentiment analysis," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, pp. 46–51.
[23] M. Kim, J. Ryu, D. Cha, and M. K. Sim, "Stock Price Prediction Using Sentiment Analysis: from

'Stock Discussion Room' in Naver," *J. Soc. E-bus. Stud.*, vol. 25, no. 4, 2022.
[24] C. S. Wu, M. Badshah, and V. Bhagwat, "Heart disease prediction using data mining techniques," *ACM Int. Conf. Proceeding Ser.*, pp. 7–11, 2019, doi: 10.1145/3352411.3352413.
[25] P. Shetgaonkar and S. Aswale, "Heart disease prediction using data mining techniques," *Int. J. Eng. Res. Technol.*, vol. 10, no. 2, pp. 281–286, 2021.
[26] L. Ji, X. Zhang, and L. Zhang, "Research on the algorithm of education data mining based on big data," in *2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI)*, 2020, pp. 344–350.
[27] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Front. Energy Res.*, vol. 9, p. 652801, 2021.
[28] E. Uncuoglu *et al.*, "Comparison of neural network, Gaussian regression, support vector machine, long short-term memory, multi-gene genetic programming, and M5 Trees methods for solving civil engineering problems," *Appl. Soft Comput.*, vol. 129, p. 109623, 2022.
[29] T. Han, L. Zhang, Z. Yin, and A. C. C. Tan, "Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine," *Measurement*, vol. 177, p. 109022, 2021.
[30] A. Afradi and A. Ebrahimabadi, "Comparison of artificial neural networks (ANN), support vector machine (SVM) and gene expression programming (GEP) approaches for predicting TBM penetration rate," *SN Appl. Sci.*, vol. 2, no. 12, pp. 1–16, 2020, doi: 10.1007/s42452-020-03767-y.
[31] M. S. Bakay and Ü. Ağbulut, "Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms," *J. Clean. Prod.*, vol. 285, p. 125324, 2021, doi: https://doi.org/10.1016/j.jclepro.2020.125324.
[32] A. Prasetyo and T. Faturohman, "Financial Distress and Financial Performance Analysis of Highway Companies Before and During the COVID-19 Pandemic: Evidence from Indonesia Stock Exchange," in *Comparative Analysis of Trade and Finance in Emerging Economies*, vol. 31, W. A. Barnett and B. S. Sergi, Eds. Emerald Publishing Limited, 2023, pp. 151–165. doi: 10.1108/S1571-038620230000031021.
[33] N. S. Utomo, S. S. Devi, and H. Siregar, "Financial Performance Analysis of Construction State-Owned Enterprises Listed in Indonesia Stock Exchange During Covid-19," *Amwaluna J. Ekon. Dan Keuang. Syariah*, vol. 6, no. 2, pp. 244–268, 2022.
[34] S. R. Ika, R. A. Nugroho, B. A. Santoso, N. F. Takril, and A. K. Widagdo, "Does the COVID-19 epidemic impact on economic sustainability of big agricultural firms in Indonesia?," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1241, no. 1, p. 12050, 2023, doi: 10.1088/1755-1315/1241/1/012050.
[35] "Stock Price Movements PT. Antam."
[36] "Stcok Price Movement PT. WIKA."
[37] "Stock Price Movements PT. Sinar Mas."
[38] K. Chen, B. Xue, M. Zhang, and F. Zhou, "An Evolutionary Multitasking-Based Feature Selection Method for High-Dimensional Classification," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 7172–7186, 2022, doi: 10.1109/TCYB.2020.3042243.
[39] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Comput. Appl.*, vol. 32, pp. 9713–9729, 2020.
[40] S. Parveen, Z. W. Satti, Q. A. Subhan, and S. Jamil, "Exploring market overreaction, investors' sentiments and investment decisions in an emerging stock market," *Borsa Istanbul Rev.*, vol. 20, no. 3, pp. 224–235, 2020.
[41] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and D. N. Rutledge, "New data preprocessing trends based on ensemble of multiple preprocessing techniques," *TrAC Trends Anal. Chem.*, vol. 132, p. 116045, 2020, doi: https://doi.org/10.1016/j.trac.2020.116045.
[42] Y. Yang *et al.*, "Clinical characteristics of hospitalized mild/moderate COVID-19 patients with a prolonged negative conversion time of SARS-CoV-2 nucleic acid detection," *BMC Infect. Dis.*, vol. 21, pp. 1–8, 2021.
[43] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey, "WeatherBench: a benchmark data set for data-driven weather forecasting," *J. Adv. Model. Earth Syst.*, vol. 12, no. 11, p. e2020MS002203, 2020.
[44] S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, "Integrated churn prediction and customer segmentation framework for telco business," *IEEE Access*, vol. 9, pp. 62118–62136, 2021.
[45] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?," *Comput. Stat.*, vol. 36, no. 3, pp. 2009–2031, 2021.

[46] H. L. Vu, K. T. W. Ng, A. Richter, and C. An, "Analysis of input set characteristics and variances on k-fold cross validation for a Recurrent Neural Network model on waste disposal rate estimation," *J. Environ. Manage.*, vol. 311, p. 114869, 2022.
[47] D.-C. Feng *et al.*, "Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach," *Constr. Build. Mater.*, vol. 230, p. 117000, 2020.