Recursive Journal of Informatics
Vol. 3, No. 2, September 2025
p-ISSN 2963-5551 https://journal.unnes.ac.id/journals/index.php/rji e-ISSN 2986-6588

Selection of Trading Indicators Using Machine Learning and Stock Close Price Prediction with the Long Short Term Memory (LSTM) Method

Alfandy Himawan Bagus Rafli[1], Aji Purwinarko[2]
[1,2]Computer Science Department, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Semarang, Indonesia

Abstract. Humans have a limit to their physical ability to work, so investment is needed to meet their needs and other goals according to their wants and needs. Investment has many types and risks according to the portion of the return value, such as mutual funds, bonds and stocks. Stocks are a form of investment that has a high risk because of the rapid fluctuations in stock values. Prediction of stock movements is usually assisted by indicators, but predictions using indicators require complex analysis because of the diverse periods and different movements in each stock data case.
Purpose: To predict the closing price of BBCA and BBRI shares in the next 10 days by considering the count of technical indicators in the form of Moving average (MA), Exponential moving average (EMA), Rate Of Change (ROC), Price Momentum, Relative Strength Index (RSI), Stochastic Oscillator in periods 21, 63 and 252.
Methods/Study design/approach: This research was conducted by comparing the accuracy of Random Forest, Decision Tree, KNN, SVM using K-fold Cross Validation then the method with the best accuracy was used to find out how much velue from the trading indicators used and predict the closing price of shares per day at BBRI and BBCA companies for the next 10 day period using the LSTM algorithm.
Result/Findings: The best accuracy in the k-fold cross validation process is random forest. random forest is used to train indicator data in determining 5 indicators along with the period that has the highest value, in this test it produces values on BBCA data in order, namely ROC63, RSI63, MOM63, MA252, EMA21 while on BBRI data in order, namely ROC63, MOM63, RSI63, MA252, MA21. This indicator is used in the price forecasting process with the LSTM method to determine the closing price in the next 10 days. The LSTM method in this study resulted in 96.8% accuracy for BBCA and 96.4% accuracy for BBRI.
Novelty/Originality/Value: The forecasting accuracy on BBCA is 96.8% and the forecasting accuracy on BBRI is 96.4%. This shows that the accuracy results are classified as good because the prediction results are close to the actual results. The data training process is expected to help traders in making stock buying and selling decisions that are adjusted to the fundamental aspects of the company.

Keywords: lstm, stock, prediction, historical data, trading, machine learning

Received January 2024 / **Revised** September 2025 / **Accepted** September 2025

INTRODUCTION

People around the world have started to realize the importance of investment [1] [2]. This investment can be in the form of stocks, deposits, property and gold. Shares are a form of proof of ownership of a company that is invested in. Stocks are an invested form of ownership of a company. There are many factors that affect stock prices [3]. Rational factors and various irrational factors determine the purchase of shares [4]. Rational factors are based on calculations that have a fundamental basis, while irrational factors are usually formed based on market trends so that the share price of a particular company soars even though in reality the company does not necessarily have a value that matches the price. Nowadays, media factors such as news articles, social media can also influence the decisions of investors [5].

Indonesia has huge potential in the stock market. But keep in mind that stocks also have the risk of loss. This is influenced by many factors such as weak company fundamentals, market volatility, government policies, rumors and market sentiment [6]. This makes the capital market very dynamic and constantly

[1]*Corresponding author.
Email addresses: alfandyhimawan@students.unnes.ac.id (Rafli)
DOI: 10.15294/rji.v3i2.945

moving. In this condition, it still brings people to invest [7]. Both by making long-term investments and short-term investments or commonly referred to as trading [8].

In predicting stock price trends, there is a level of vulnerability due to many factors that already exist so that more measurement tools are needed in maximizing profits and minimizing capital losses. Among the methodologies for predicting stock trend movements, the things that need to be considered are Engineering Analysis [9], Time Series Forcasting [10], Machine learning and Data Mining [11], and Modeling and predicting stock volatility using differential equations [12].

Stocks are influenced by Fundamental Analysis which focuses on the economic and financial factors of a company, and Technical Analysis which focuses on stock price movements through historical data on securities [13]. This technical analysis has many types and will continue to evolve and come up with new ways of calculation. Widely recognized technical analysis such as Moving Average (MA), Rate Of Change (ROC), Price Momentum, Relative Strength Index (RSI) and Stochastic Oscillator [14].

Machine learning method is a type of algorithm method that is included in the Artificial Neural Network (ANN). This network is a network formed from neurons in which there are three layers: input layer, hidden layer, output layer [15]. However, the problem experienced by ANN is that it is less effective when neural networks that have many hidden layers experience vanishing gradients so that ANN is unable to capture sequential information on the required input data. Recurrent Neural Network (RNN) overcomes the problem with the ability to store more and longer networks, resulting in more optimal output [16]. One RNN method that is often used to predict a sequence of numbers and time-based data information is the Long Short Term Memory (LSTM) method. The LSTM algorithm is designed to fulfill the shortcomings of ANN while still maintaining the advantages of predicting more accurate information [17].

METHODS

In this study, the calculation of trading indicators with periods of 21, 63 and 252 then selected the KNN, SVM, Decision Tree and Random forest methods with K-Fold Cross Validation. The highest method accuracy is used to process 5 indicators with the highest velue importance. The selected indicators are used to forecast the closing prices of shares in BBCA and BBRI for the next 10 days using the LSTM method. This research is divided into four main steps, as described below.

Data Research

stocks of PT Bank Rakyat Indonesia (Persero) Tbk and PT Bank Central Asia Tbk with a data period of January 2019 - January 2024 and not suspended. Data obtained from Yahoo Finance contains historical data per day. Predictions from the stock price chart are carried out only for the next 10 days, this prediction has 3 possibilities, namely an increase in price, the price will remain, or there will be a price decline. Trials were conducted with 3 data divisions, namely 70% train data - 30% test data, 80% train data - 20% test data, and 90% train data - 10% test data. The amount of data can be summarized in table 1.

Table 1. Data sharing

Data Train - Data Test	BBCA and BBRI data
70% - 30%	853 - 366 data
80% - 20%	975 – 244 data
90% - 10%	1097 – 122 data

Pre-processing

In this stage, the dataset is checked if there are missing values and data that are repeated or irrelevant to use. This process is carried out twice, namely at the beginning of calling the data and when the data has been calculated for the indicator results. This is done to make efficiency so as to reduce computation time. The data cleaning process is depicted in Figure 1.

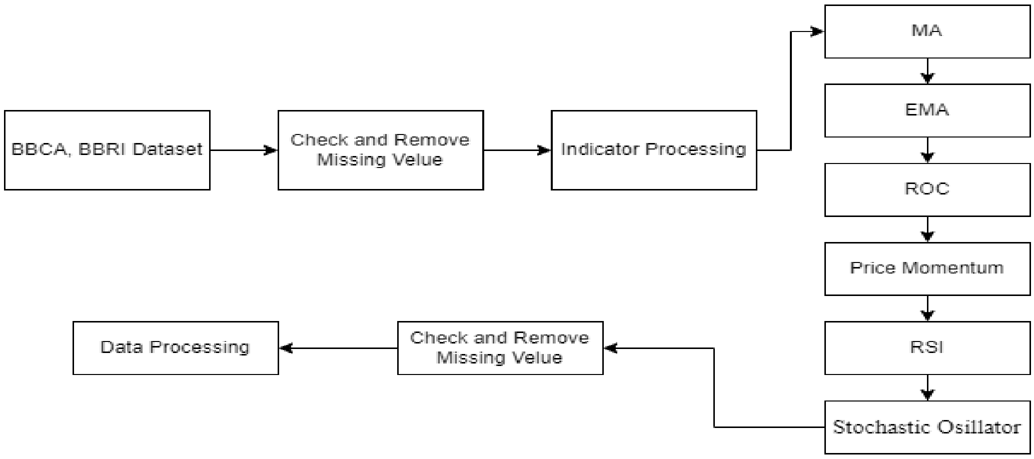

Figure 1. Pre-processing data

Train data

The model produced in this stage will be used to conduct trials at the Testing Data stage. Where all information or data from training results are collected in 1file called lstm_weights.h5. The flowchart of our proposed method is shown in Figure 2.

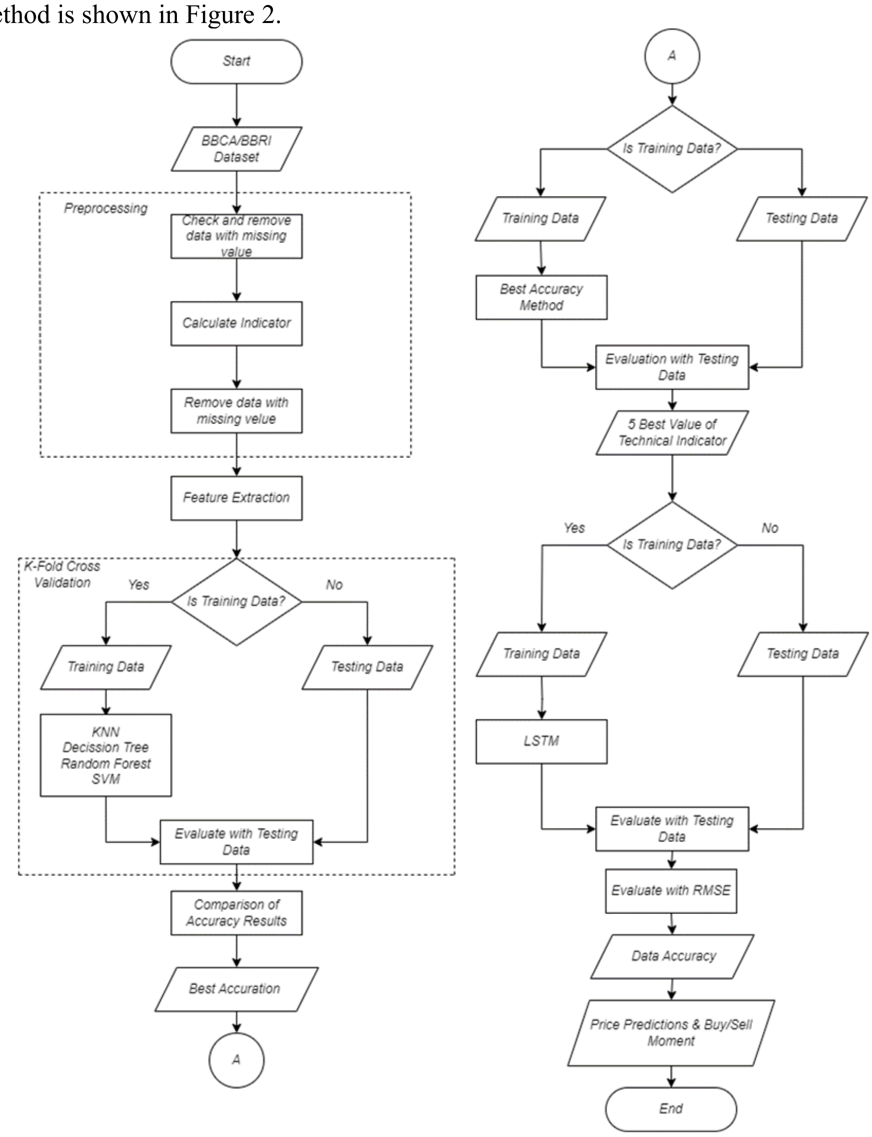

Figure 2. Research stages

Model Evaluation

In testing the effectiveness and accuracy of a prediction result obtained from a machine learning model, it is tested through an evaluation matrix. The evaluation matrix is used as a benchmark for the accuracy of a prediction result with actual results [18]. Commonly used evaluation matrices include Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

RMSE

RMSE is the square root of the average squared difference between the actual value and the resulting predicted value [19]. The smaller the RMSE value, the better the quality of the resulting prediction. The calculation of RMSE can be calculated through Equation 1 [20].

$$RMSE = \sqrt{\sum \frac{(Y' - Y)^2}{n}} \tag{1}$$

MAPE

MAPE is a very commonly used evaluation method, this is because the results of MAPE are in the form of percentages so it is easier to understand and compare [21]. MAPE is calculated by calculating the average of the percentage difference between predicted and actual values [22]. MAPE is a calculation of the percentage error of a prediction result compared to the actual value, so the smaller the MAPE value, the better the quality of the resulting prediction data. The MAPE calculation can be defined by Equation 2 [23].

$$MAPE = \frac{\sum_{t=1}^{n} \left| \left(\frac{At - Ft}{At} \right) 100 \right|}{n} \tag{2}$$

RESULT AND DISCUSSION

Pre-processing Process

At this stage, it is done to call the data and select the data needed. The preprocessing stage is used to eliminate data that has missing values. Data checking is also carried out on data types such as in the date column detected to have the data type "object" as in Figure 1 changed to "datetime" as in Figure 2 this stage is carried out to facilitate the next process.

```
[3] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1220 entries, 0 to 1219
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Date       1220 non-null   object
 1   Open       1220 non-null   float64
 2   High       1220 non-null   float64
 3   Low        1220 non-null   float64
 4   Close      1220 non-null   float64
 5   Adj Close  1220 non-null   float64
 6   Volume     1220 non-null   int64
dtypes: float64(5), int64(1), object(1)
memory usage: 66.8+ KB
```

Figure 3. Date column with object datatype

```
[5] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1220 entries, 0 to 1219
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Date       1220 non-null   datetime64[ns]
 1   Open       1220 non-null   float64
 2   High       1220 non-null   float64
 3   Low        1220 non-null   float64
 4   Close      1220 non-null   float64
 5   Adj Close  1220 non-null   float64
 6   Volume     1220 non-null   int64
dtypes: datetime64[ns](1), float64(5), int64(1)
memory usage: 66.8 KB
```

Figure 4. Date column with datetime datatype

Data Retrieval

K-Fold Cross Validation

The k-fold cross validation process is used in determining the accuracy of each data division on BBCA and BBRI data which can be seen in the table 2 and table 3.

Table 2. BBCA K-Fold Cross Validation accuracy results

Data train – data test	KNN	*Decision tree*	SVM	*Random forest*
70% - 30%	0.952157	0.932392	0.940196	0.972118
80% - 20%	0.956473	0.928675	0.949456	0.979099
90% - 10%	0.951899	0.941202	0.951899	0.978365

Table 3. BBRI K-Fold Cross Validation accuracy results

Data train – data test	KNN	Decision tree	SVM	Random forest
70% - 30%	0.968078	0.970078	0.964118	0.976978
80% - 20%	0.966818	0.966788	0.968542	0.979008
90% - 10%	0.978293	0.973630	0.970529	0.986058

Indicator Accuracy

Random forest shows the highest accuracy results from other methods. Random forest is used to determine the velue importance of indicators with each period. The velue importance results can be seen in Figure 3 and Figure 4.

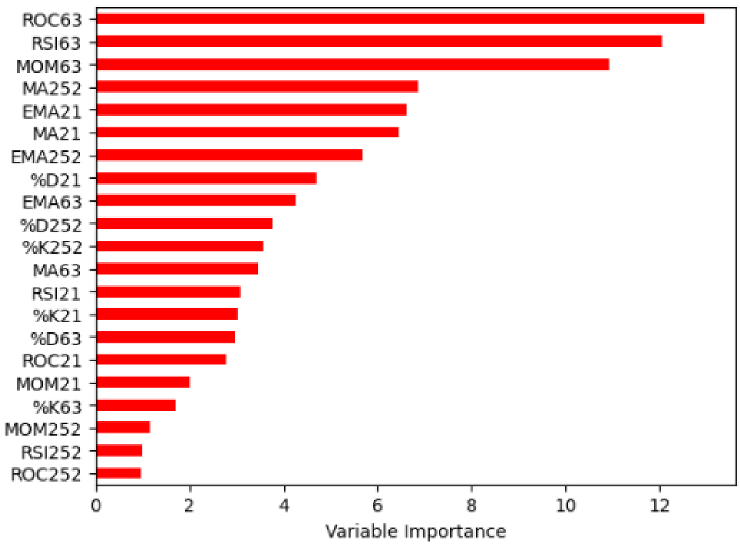

Figure 5. BBCA Indicator accuracy results

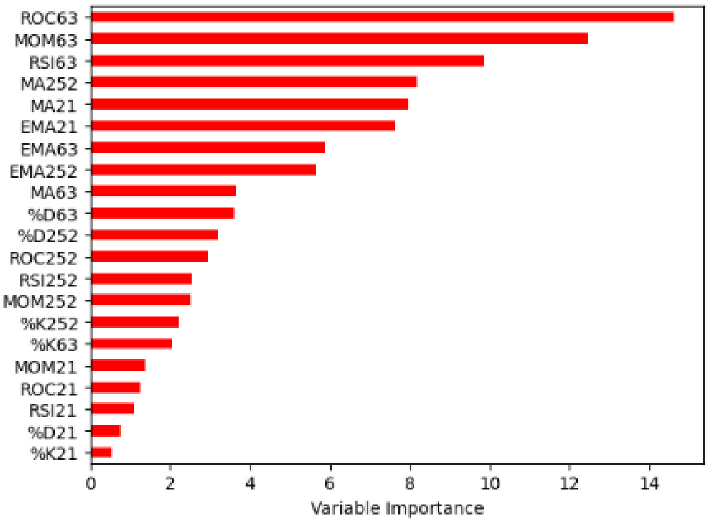

Figure 6. BBRI Indicator accuracy results

Model Evaluation

The actual data starts with the closing price data on January 4, 2024 until 10 days in the future when the capital market is active. Each table consists of observation values or actual values that occur which are compared with the predicted values generated according to the calculation method. Each table is divided into 3 data model divisions, each of which is 70% data train 30% data test (70%-30%), 80% data train 20% data test (80%-20%), and 90% data train 10% data test (90%-10%). The results of the RMSE and MAPE evaluation matrices are shown in Table 4.

Table 4. Accuracy result

Pembagian data	RMSE	MAPE - Accuracy
BBCA 70% - 30%	418,2	4,3% - 95,7%
BBCA 80% - 20%	305,6	3,2% - 96,8%
BBCA 90% - 10%	518,2	5,4% - 94,6%
BBRI 70% - 30%	349,4	5,9% - 94,1%
BBRI 80% – 20%	219	3,6% - 96,4%
BBRI 90% - 10%	407	6,9% - 93,1%

Discussion

This research uses historical price datasets on the stock prices of BBCA and BBRI Companies with data periods January 2019 - January 2024. This data consists of date, open, high, low, close, adj close, volume. This dataset can be accessed openly on the Yahoo Finannce platform which provides a variety of historical stock price data around the world. The writing of this research was inspired by personal anxiety about future financial needs and simplifying the concept of old prediction methods that are still calculated manually modified into machine learning. This concept can be used for various transaction service providers on stocks as a predictive analysis.

The dataset used in this research contains information about prices in detail which is classified per day on each existing data. In implementing machine learning, it is necessary to check the data to be used, so data preprocessing is carried out to check the data and change the type of data accordingly so that it can be processed. Data checking is done to look for data that is NaN or missing value then the data is deleted so that all data can be completely processed. Changing the data type is done on the date data which is defined as an object then changed in the form of a datetime. Data that has gone through the preprocessing stage is used to calculate technical indicators in the form of MA, EMA, ROC, price momentum, RSI and Stochastic Oscillator with the periods used are 21, 63, 252 respectively. This period serves as a grouping of the data used which is useful in finding results according to the calculation equation for each indicator.

This research applies filtration to several types of machine learning methods such as decision tree, random forest, KNN, and SVM with K-Fold Cross Validation so that it will produce the accuracy level of each method for training technical indicator data previously calculated by data division trials of 70% train data - 30% test data, 80% train data - 20% test data, 90% train data - 10% test data which results in random forest as a method that has the highest accuracy in each data division. Random forest is used to train technical indicator data so as to produce signals to determine prices. This data is then used to train data using LSTM to predict the closing price in the next 10 days. The application of LSTM is also done by dividing the data with a proportion of 70% train data - 30% test data, 80% train data - 20% test data, 90% train data - 10% test data. Data evaluation is carried out with the RMSE and MAPE evaluation matrices. A recap of the results of the evaluation matrix can be seen in Table 4.

The accuracy value of the RMSE and MAPE evaluation matrix on the 80% - 20% data division shows better results than other data division experiments. The results of this evaluation matrix show relatively good results because the predicted value produces a stock closing price that is not far from the actual price. Manuscripts can be presented with the support of tables, graphs or images which needed to clarify the results of presentation verbally. Results and discussion is shown clearly and concisely.

CONCLUSION

In this study it can be concluded that in the machine learning selection process using the K-Fold Cross Validation method, the highest value result is the random forest method. The random forest algorithm has the highest accuracy and value when compared to the decision tree algorithm, KNN, and SVM on the two data tested. random forest is used to train indicator data in determining 5 indicators along with the period that has the highest value, in this test it produces values on BBCA data in order, namely ROC63, RSI63, MOM63, MA252, EMA21 while on BBRI data in order, namely ROC63, MOM63, RSI63, MA252, MA21. Each period indicator has relatively the same results, it's just different in the 5th order on BBCA resulting in EMA21 while on BBRI MA21. This indicator is used in the price forecasting process with the LSTM method to determine the closing price for the next 10 days. The LSTM method in this study resulted in an accuracy value of 96.8% for BBCA and 96.4% accuracy for BBRI. With this research, it is hoped that it can be developed by using data on company shares and other sectors and deepening indicators and more periods so that it can be explored more deeply in the algorithms used and allow maximum accuracy results.

REFERENCES

[1] J.-L. Rodríguez-Sánchez, T. González-Torres, A. Montero-Navarro, and R. Gallego-Losada, "Investing time and resources for work–life balance: The effect on talent retention," *Int. J. Environ. Res. Public Health*, vol. 17, no. 6, p. 1920, 2020, doi: https://doi.org/10.3390/ijerph17061920.

[2] O. S. Sabirov, B. T. Berdiyarov, A. S. Yusupov, A. T. Absalamov, and A. I. U. Berdibekov, "Improving Ways to Increase the Attitude of the Investment Environment," *Rev. Geintec-Gestao Inov. E Tecnol.*, vol. 11, no. 2, pp. 1961–1975, 2021.

[3] D. T. N. Huy, P. M. Dat, and P. T. Anh, "BUILDING AN ECONOMETRIC MODEL OF SELECTED FACTORS'IMPACT ON STOCK PRICE: A CASE STUDY.," *J. Secur. Sustain. Issues*, vol. 9, 2020, doi: 10.9770/jssi.2020.9.M(7).

[4] M. Hani'ah, M. Z. Abdullah, W. I. Sabilla, S. Akbar, and D. R. Shafara, "Google Trends and Technical Indicator based Machine Learning for Stock Market Prediction," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 22, no. 2, pp. 271–284, 2023, doi: https://doi.org/10.30812/matrik.v22i2.2287.

[5] M. G. Pradana, A. C. Nurcahyo, and P. H. Saputro, "Pengaruh Sentimen Di Sosial Media Dengan Harga Saham Perusahaan," *J. Ilm. Edutic Pendidik. dan Inform.*, vol. 6, no. 2, pp. 67–77, 2020, doi: 10.21107/edutic.v6i2.6992.

[6] G. D. Abuselidze and A. N. Slobodianyk, "Value Assessment of shares of corporate issuers by applying the methods of fundamental analysis in the stock exchange market," in *The Challenge of Sustainability in Agricultural Systems: Volume 2*, Springer, 2021, pp. 25–39. doi: https://doi.org/10.1007/978-3-030-72110-7_3.

[7] M. Hu, D. Zhang, Q. Ji, and L. Wei, "Macro factors and the realized volatility of commodities: a dynamic network analysis," *Resour. Policy*, vol. 68, p. 101813, 2020, doi: https://doi.org/10.1016/j.resourpol.2020.101813.

[8] A. Roberts, H. Choer Moraes, and V. Ferguson, "Toward a geoeconomic order in international trade and investment," *J. Int. Econ. Law*, vol. 22, no. 4, pp. 655–676, 2019, doi: https://doi.org/10.1093/jiel/jgz036.

[9] U. Safryani, A. Aziz, and N. Triwahyuningtyas, "Analisis Literasi Keuangan, Perilaku Keuangan,

Dan Pendapatan Terhadap Keputusan Investasi," *J. Ilm. Akunt. Kesatuan*, vol. 8, no. 3, pp. 319–332, 2020.
[10] B. Lim and S. Zohren, "Time-series forecasting with deep learning: a survey," *Philos. Trans. R. Soc. A*, vol. 379, no. 2194, p. 20200209, 2021, doi: https://doi.org/10.48550/arXiv.2004.13408.
[11] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Syst. Appl.*, vol. 166, p. 114060, 2021, doi: https://doi.org/10.1016/j.eswa.2020.114060.
[12] C. Rackauckas *et al.*, "Universal differential equations for scientific machine learning," *arXiv Prepr. arXiv2001.04385*, 2020.
[13] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," *Artif. Intell. Rev.*, vol. 53, no. 4, pp. 3007–3057, 2020, doi: 10.1007/s10462-019-09754-z.
[14] T. Stein, "Forecasting the equity premium with frequency-decomposed technical indicators," *Int. J. Forecast.*, vol. 40, no. 1, pp. 6–28, 2024, doi: https://doi.org/10.1016/j.ijforecast.2022.12.001.
[15] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, no. 11, p. e00938, 2018, doi: https://doi.org/10.1016/j.heliyon.2018.e00938.
[16] R. Cui, M. Liu, and A. D. N. Initiative, "RNN-based longitudinal analysis for diagnosis of Alzheimer's disease," *Comput. Med. Imaging Graph.*, vol. 73, pp. 1–10, 2019, doi: https://doi.org/10.1016/j.compmedimag.2019.01.005.
[17] J. Gonzalez and W. Yu, "Non-linear system modeling using LSTM neural networks," *IFAC-PapersOnLine*, vol. 51, no. 13, pp. 485–489, 2018, doi: https://doi.org/10.1016/j.ifacol.2018.07.326.
[18] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, 2021, doi: 0.7717/peerj-cs.623.
[19] J. Rasheed, A. Jamil, A. A. Hameed, M. Ilyas, A. Özyavaş, and N. Ajlouni, "Improving stock prediction accuracy using cnn and lstm," *2020 Int. Conf. Data Anal. Bus. Ind. W. Towar. a Sustain. Econ.*, pp. 1–5, 2020, doi: 10.1109/ICDABI51230.2020.9325597.
[20] M. Ćalasan, S. H. E. A. Aleem, and A. F. Zobaa, "On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: A novel exact analytical solution based on Lambert W function," *Energy Convers. Manag.*, vol. 210, p. 112716, 2020, doi: https://doi.org/10.56855/analysis.v1i1.177.
[21] P. Borlido, J. Schmidt, A. W. Huran, F. Tran, M. A. L. Marques, and S. Botti, "Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning," *npj Comput. Mater.*, vol. 6, no. 1, p. 96, 2020, doi: https://doi.org/10.1038/s41524-020-00360-0.
[22] S. Muzaffar and A. Afshari, "Short-term load forecasts using LSTM networks," *Energy Procedia*, vol. 158, pp. 2922–2927, 2019, doi: https://doi.org/10.1016/j.egypro.2019.01.952.
[23] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "Stock closing price prediction using machine learning techniques," *Procedia Comput. Sci.*, vol. 167, pp. 599–606, 2020, doi: https://doi.org/10.1016/j.procs.2020.03.326.