

Support Vector Machine for Predicting Candlestick Chart Movement on Foreign Exchange

Annisa Nurul Puteri¹, Suryadi Syamsu¹, Topan Leoni Putra¹, Andita Dani Achmad²

¹Universitas Teknologi Akba Makassar, Makassar, Indonesia

²Universitas Fajar, Makassar, Indonesia

Article Info

Article history:

Received January 10, 2023

Revised March 06, 2023

Accepted March 15, 2023

Keywords:

Candlestick Chart
Classification

Foreign Exchange
Machine Learning

Prediction

Support Vector Machine

ABSTRACT

Foreign Exchange, commonly called Forex, is a form of investment in the non-real sector in great demand. Forex is a marketplace that specializes in foreign exchange trading. Technology advancements have made it easy to monitor investment conditions in real time and present them in an easy-to-understand graphical form. As a result, predictions are closely related to investment, starting from market sentiment and economic conditions to technical matters. One of the Artificial Intelligence methods that can be used in classifying is the Support Vector Machine (SVM). SVM is a machine learning classification method based on the Structural Risk Minimization (SRM) principle to find the best hyperplane that separates two classes in the input space that determines the classification decision function by minimizing empirical risk. This study used candlestick patterns to predict foreign exchange chart movements using the Support Vector Machine (SVM) classification method. The purpose of this study was to measure the accuracy of the Support Vector Machine method in making predictions using candlestick patterns so that it can assist traders in making decisions in forex trading. The accuracy level obtained from the data classification results reached 90.72% with a precision of 87.69%. With a relatively good level of accuracy, the Support Vector Machine (SVM) method can be used to predict chart movements in foreign exchange using candlesticks to indicate the current trend's direction.

Copyright ©2022 The Authors

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Suryadi Syamsu, +6285255290904,
Program Studi Teknik Informatika,
Universitas Teknologi Akba Makassar, Makassar, Indonesia,
Email: adi@akba.ac.id

How to Cite: A. Puteri, S. Syamsu, T. Putra, and A. Achmad, "Support Vector Machine for Predicting Candlestick Chart Movement on Foreign Exchange", *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, no. 2, pp. 249-260, Mar. 2023.

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

1. INTRODUCTION

Investment in the non-real sector has recently started to become a concern. Many vendors have started to appear to offer investment forms that are expected to attract investors with high or low financial ability. Foreign Exchange, commonly called Forex, is a form of non-real sector investment in great demand. Forex is a marketplace specialized in foreign exchange trading. With the main characteristics of high-risk, high-return, Forex offers an investment with high risk. However, it will provide extraordinary benefits if investors can manage risk and make good predictions [1, 2]. On the other hand, the inability of investors to manage risk can result in significant losses. Therefore, investors need to analyze indications of the direction of the current trend in order to minimize the risk of possible losses.

Existing technological advances have made real-time conditions of the investment environment easy to monitor and presented in easy-to-understand graphical form. One example is the MetaTrader 4 application which displays the changing currency exchange conditions as a candlestick chart. The candlestick chart has many functional visual patterns and can interpret prices during the session simply so that it is easy to explain in decision-making. As a result, the candlestick chart is one of the most widely used charts [3, 4].

Predictions are closely related to investment, starting from market sentiment, economic conditions, and technical matters. These indicators have been widely used to predict when and what to do as an investor. Investors need accurate predictions in making decisions to reduce the risks that can occur in their investments. In the Forex prediction process, candlestick charts can make it easier for users to extract the information they need. Predictions can be made using Machine Learning or Data Mining techniques. Data mining techniques can produce added value from a data set in the form of previously unknown knowledge, such as information about relationships between variables by predicting patterns, identifying characteristics, and classifying data [5, 6]. Support Vector Machine is a data mining technique that can predict accurately [7]. According to L. Auria and R. A. Moro, cited in [8], Support Vector Machine is widely used in various studies to perform automatic classification, including image recognition, medical analysis, or to make predictions. The SVM algorithm is based on statistical theory and gives better results than other methods because it works effectively when removing irrelevant features [9].

Research [10, 11] in 2020 created a prediction model for foreign currency exchange rates using last year's data as input. Three machine learning models are used to compare their performance in predicting foreign currency exchange rates, namely Long Short Term Memory (LSTM), Simple Recurrent Neural Network (SRNN), and Gated Recurrent Unit (GRU) in predicting currency exchange rates of 22 countries against the United States. Stated Dollar (USD) simultaneously. The best results were obtained from the LSTM Model with 200-100-30 coating units. Similar results are shown with the three activation functions; Linear, Sigmoid, and Tanh. This research indicates that Neural Networks can predict many foreign exchange rates simultaneously very efficiently.

Stock price movements were predicted by [4] in 2021. This study proposes a Deep Predictor for Price Movement (DPP) using candlestick charts on the proposed stock history data. First, DPP decomposes the given candlestick chart into several sub-charts. Then, the best representation of the sub-charts is obtained using the CNN autoencoder, and GRU makes price movement predictions. The DPP-based model performance assessment uses trading data from the Taiwan Stock Exchange Capitalization-Weighted Stock Index and the stock market index, Nikkei 225, for the Tokyo Stock Exchange. The experimental results show that DPP is superior to other methods.

Predicting foreign exchange was carried out by [12] in 2021 using Digital Signal Processing. A digital processing model was developed to predict foreign exchange using the ARIMA algorithm and Artificial Neural Network. The data ranges from 20 years with five currencies; USD, Swiss Pound, Yen, Euro, and Franc. This study produces a minimum error value of 0.7 within 26 seconds, measured using the Root Mean Square Error (RMSE).

M.S. Islam and E. Hossain conducted research [13] in 2021 that combined Gated Recurrent Unit (GRU) and Long Short Term Memory (LSTM) to predict the future closing price of FOREX currencies. The currency data used in this study are EUR/USD, GBP/USD, USD/CAD, and USD/CHF. Furthermore, model performance is validated using MSE, RMSE, MAE, and R2. The results obtained are the performance of the GRU-LSTM combined model, giving the best results for the GBP/USD and USD/CAD currency pairs based on the performance values of MSE, RMSE, and MAE. Meanwhile, regarding the performance of the R2 system built, M.S. Islam and E. Hossain show the best performance of all systems that have been compared and prove that the system built is a system model with the least risk among all models.

In 2022, a candlestick pattern prediction ability test was carried out by [14] to determine trading positions. The Gramian Angular Field (GAF) presents candlestick patterns as images to identify 3 hours and 5 hours of 6 candlestick patterns using a Convolutional Neural Network (CNN). Predictions use Long Short-Term Memory (LSTM) to predict the closing price. This study shows that CNN successfully detected 3-hour and 5-hour candlestick patterns with an accuracy rate of 90% and 93%. LSTM can predict closing price trends with an RMSE score of 155.458 and a MAPE of 0.9754% with a 10-hour look-back display. The holding duration is 3 hours, and CNN-LSTM as an additional model produces 85 candlestick patterns from the test data that can be recognized with an accuracy of up to 82.7%, better than candlestick pattern recognition using CNN alone, which achieves an accuracy

of only 60%. The CNN-LSTM combination can increase the predictive power of candlestick patterns and offers more advantages in determining trading positions.

Based on the research that has been done, where candlesticks can be used to indicate the current trend direction using Convolutional Neural Network (CNN) method and Long Short-Term Memory (LSTM), this study uses candlestick patterns to predict chart motion on foreign exchange with the Support Vector Machine (SVM) classification method. Furthermore, this study aims to measure the accuracy of the Support Vector Machine method in making predictions using candlestick patterns in the MetaTrader 4 application so that it can help traders make decisions in foreign exchange trading.

2. RESEARCH METHOD

This research went through several systematic stages, from data collection to obtaining foreign exchange prediction results, as seen in Figure 1. This stage consists of data retrieval, literature study, data preprocessing, data processing, system testing, and analysis.

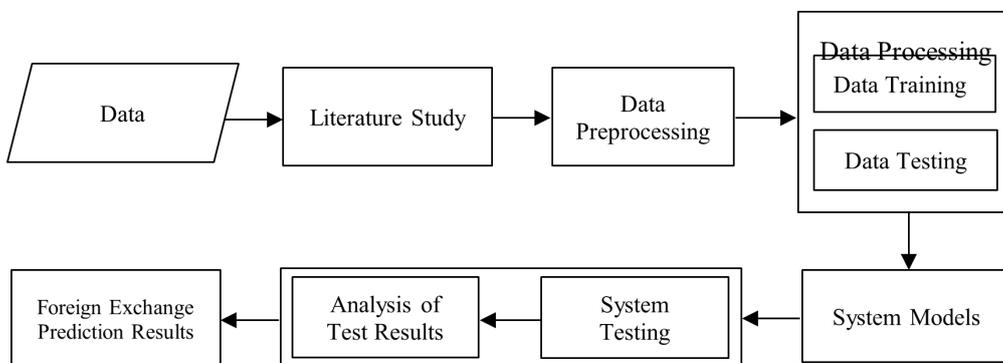


Figure 1. Research stages

2.1. Dataset

The data used in this study is a collection of candlestick patterns from 2017 to 2021, which can be accessed freely on the MetaTrader 4 application. The data consists of 7 attributes which are all numerical, with a total of 1835 data before going through the preprocessing stage. This study uses the currency pair Great Britain Pound Sterling and United States Dollar (GBP/USD) with a timeline of four hours (4h).

2.2. Preprocessing

This stage requires the process of cleaning data from missing values because it can negatively affect the performance of the method used. Irrelevant data is also cleaned at this stage [15]. At the preprocessing stage, normalization is also carried out to remove redundant data, repetitive data, and outliers and change the range of data values for each attribute to 0 to 1, not to reduce the data testing accuracy potentially (see equation (1)). Data from the MetaTrader 4 application will be selected based on the shape of the candlestick and then converted into data with three attributes, namely body, top wick, and bottom wick.

$$Normalized(X) = \frac{minRange(X - minvalue)(maxrange - minrange)}{maxValue - minValue} \quad (1)$$

Normalization in this study uses Min-Max Normalization. According to T. T. Hanifa and S. Al-Faraby in Nasution et al. [16], Min-Max Normalization is a normalization method that produces a balance of comparison values between the data before and after processing, carrying out a linear transformation of the original data. The formula for determining the normalization value can be seen in equation (1).

2.3. Classification

Classification is a data mining technique that groups data based on the closeness or attachment of the data to the sample data. A model can be found from a data set consisting of several features and classes of various other features [17]. In [18], Derek and Schnyer stated that most of the existing guided classification methods are based on traditional statistics, which can provide ideal results if the sample size is not limited. However, in practice, the number of samples obtained is limited.

Classification can be done using several methods, one of which is the Support Vector Machine (SVM). SVM is a machine learning classification method that works based on the Structural Risk Minimization (SRM) principle to find the best hyperplane that separates two classes in the input space that determines the classification decision function by minimizing empirical risk [19, 20].

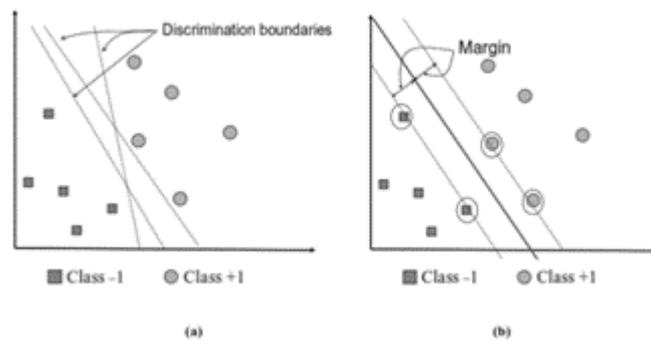


Figure 2. Defining hyperplane

Several patterns belonging to two classes: +1 and 1, are shown in Figure 2. Patterns belonging to class 1 are denoted by a square shape, while a circle shape denotes patterns in class +1. The classification problem can be interpreted by finding a line (hyperplane) separating the two groups [21]. Hyperplane determination can also be formulated using equation (2):

$$R = \frac{1}{l} \sum_{i=1}^l |(x_i) - y_i| \quad (2)$$

In Equation (2), l and f represent the sample size and classification decision function. Each SVM aims to determine the dividing hyperplane in the optimal classification class to generate low errors. However, in some cases, the classification decision function has problems in linear separation. A hyperplane is needed to determine the minimum boundary with Equation (3).

$$f_{w,b} = \sin(wx - b) \quad (3)$$

The minimization problem can be modified to allow data points to be misclassified with Equation (4) for cases that cannot be separated linearly. SVM can be applied to multiclass classification by combining SVM.

$$\begin{aligned} \text{Min} : & \frac{1}{2} w^T \\ & y_1(w \cdot x_i + b) \geq 1 \end{aligned} \quad (4)$$

2.4. Performance Evaluation

Performance evaluation is carried out to see the error rate of the classification method. The performance of the model that has been made is measured using a confusion matrix. Table 1 shows the classification results represented as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [14].

Table 1. Confusion matrix

Prediction \ Actual	TRUE	FALSE
	TRUE	TP
FALSE	FN	TN

Description:

- True Positive (TP) : positive data that is properly classified as positive data
- True Negative (TN) : negative data that is correctly classified as negative data
- False Positive (FP) : negative data are classified as positive data
- False Negative (FN) : positive data that is classified as negative data

Accuracy is the degree of closeness between actual data and predictive data. Accuracy can also be defined as the ratio of the data amount correctly classified by the system [22, 23]. The level of accuracy can be calculated using the formula (5). The level of accuracy between the answers given by the system and the requested information is called precision. Precision calculations can be done with the equation formula (6).

$$Accuracy = \frac{TP + N}{TN + FP + FN + TP} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

3. RESULT AND ANALYSIS

Data taken from the Metatrader 4 application is then processed using Matlab 2016a. After going through the preprocessing stage, the amount of data that is feasible to use is 970 data with three numeric attributes. The data is divided into training and test data, with a percentage of 70% and 30%.

3.1. Preprocessing

Data from MetaTrader 4 is processed into data first that represents the candlestick shape. In Figure 3, data will be selected based on the candlestick shape and converted into three attributes: body, top, and bottom wick. This stage is carried out to find the difference between the opening price, the highest price, the lowest price, and the closing price in one sequence.

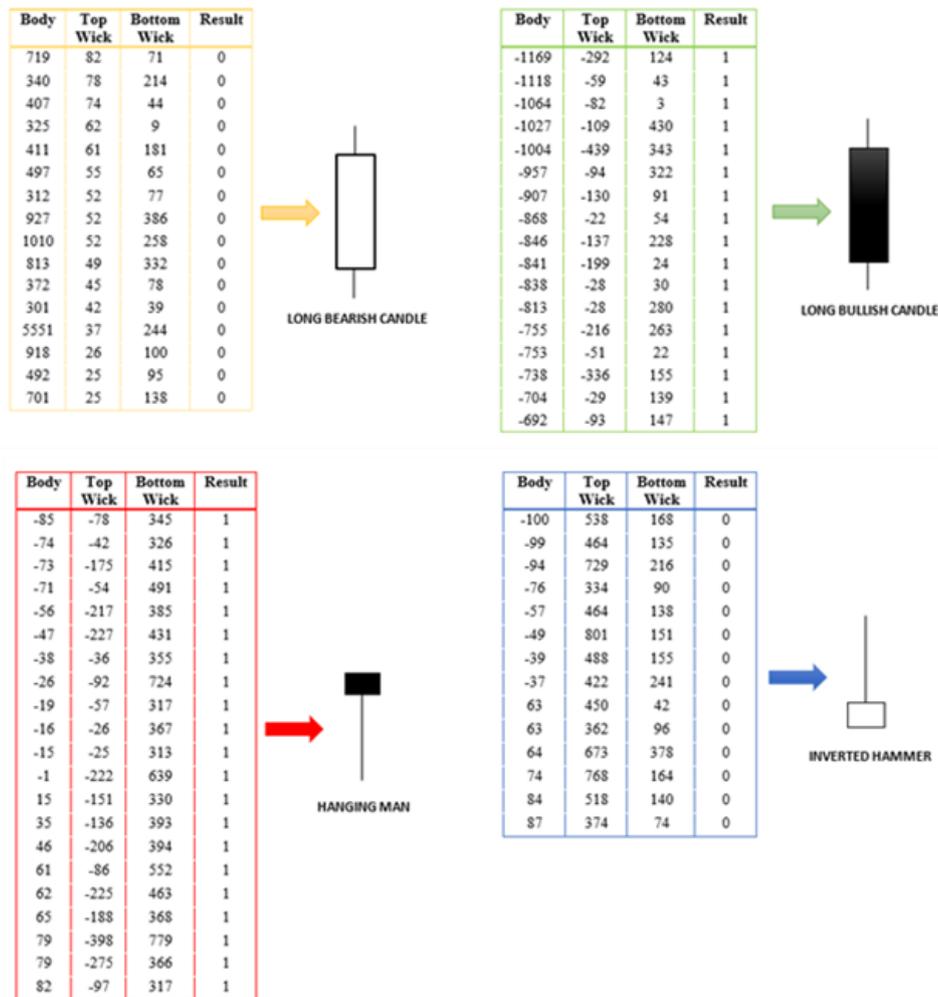


Figure 3. Candlestick data transformation

The results of the candlestick representation in the form of attributes that will be used in this study can be seen in Table 2. There are five attributes resulting from the representation of the candlestick, namely Body Candle, Top Wick Candle (full), Top Wick Candle (empty), Bottom Wick Candle (full), and Bottom Wick Candle (empty). Body Candle is described as the closing price minus the opening price, Top Wick Candle (full) is described as the highest price minus the opening price, Top Wick Candle (empty) is described as the highest price minus the closing price, Bottom Wick Candle (full) is described as the lowest price minus the opening price, and Bottom Wick Candle (empty) as the lowest price minus the closing price.

Table 2. Attribute information

No.	Attribute	Description
1	Body Candle	The closing price minus the opening price
2	Top Wick Candle (full)	The highest price minus the opening price
3	Top Wick Candle (empty)	The highest price minus the closing price
4	Bottom Wick Candle (full)	The lowest price minus the opening price
5	Bottom Wick Candle (empty)	The lowest price minus the closing price

Normalization is carried out to change the range of values for each attribute to a range of 0 to 1 using Equation (1). An example of data before and after going through the normalization stage can be seen in Table 3. In addition, the target type is also changed to 0 for the "No" class and 1 for the "Yes" class.

3.3. Analysis

Several experiments were carried out in this study to get the best performance from SVM. Figure 5 is a training visualization model to obtain optimal hyperparameters and minimize errors during cross-validation.

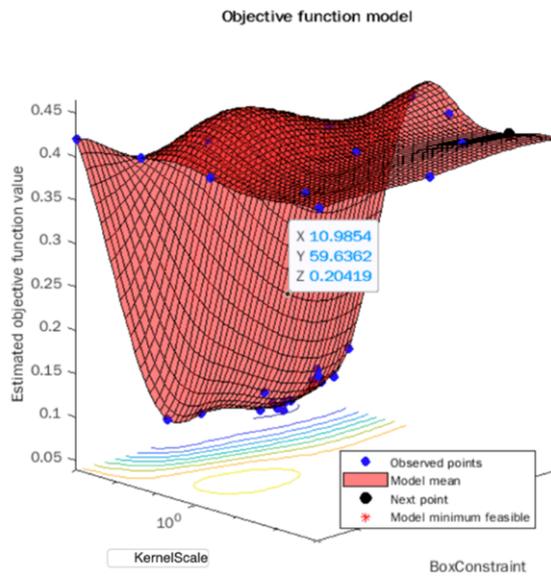


Figure 5. SVM training results in 3D model

Figure 5 is a training visualization model to obtain optimal hyperparameters and minimize errors during cross-validation. The fitsvm function will help to get optimal box constraints and kernel functions so that the observation points obtained can be more accurate. It can be seen in Figure 5 that the best observation points were obtained at X with a value of 10.9854, Y with a value of 59.6362, and Z with a value of 0.20419. A 2D visualization of hyperplane healing efforts from the input training data can be seen in Figure 6.

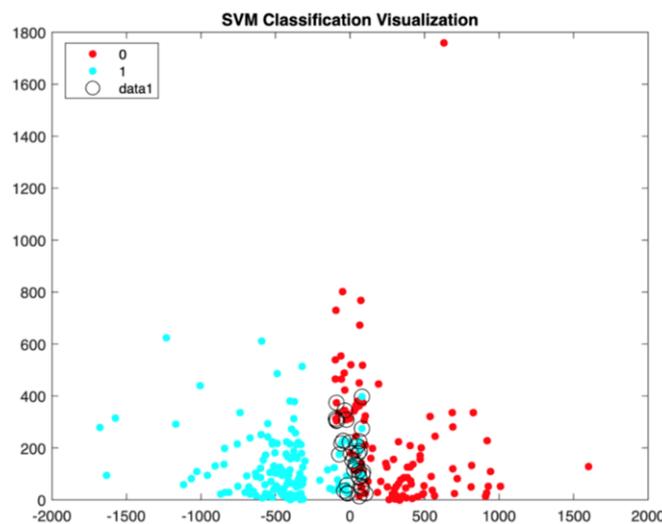


Figure 6. 2D SVM-trained hyperplane search model

Performance evaluation uses a connection matrix to see the accuracy and precision of the data being tested. The results of the classification can be seen in Table 4.

Table 4. Classification result

Prediction \ Actual	Actual	
	TRUE	FALSE
TRUE	114	16
FALSE	11	150

From the classification results in Table 4, calculations can be made to see accuracy with Equation (5) and precision with equation (6). The classification results have an accuracy rate of 90.72% and a precision of 87.69%. In this section, the researcher compared the accuracy results between the SVM method that was carried out by researchers with CNN and CNN-LSTM, which was carried out in the study [14], which can be seen in Figure 7.

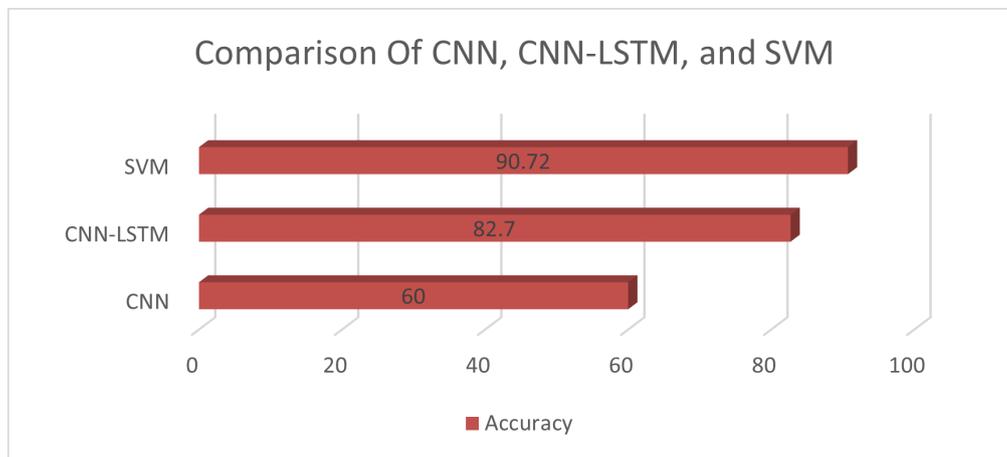


Figure 7. Comparison accuracy of the methods

It can be seen in Figure 7 that SVM has a better accuracy rate than CNN, with an accuracy rate of 60%, and the combination of CNN and LSTM, with an accuracy rate of 82.7%, was carried out by [14].

4. CONCLUSION

The accuracy level obtained from the data classification results is 90.7%, with a precision of 87.69%. This value is obtained from testing 291 data using three attributes, namely body, top wick, and bottom wick. Therefore, the Support Vector Machine (SVM) method can be used to predict the movement of charts in foreign exchange using candlesticks to indicate the current trend's direction so that it can help traders make decisions in foreign exchange trading. SVM can also be used to create system models for classification. Further research can be carried out by adding a feature selection method to achieve better accuracy.

5. ACKNOWLEDGEMENTS

The authors would like to thank Universitas Teknologi Akba Makassar, which has provided infrastructure support during the research.

6. DECLARATIONS

AUTHOR CONTRIBUTION

The authors of the first, second, and third carried out the idea and research design. The first and third authors performed data collection and analysis. The second and third authors carried out experiments and testsmanuscript writing by the first and fourth authors. All authors carried out revision and finalization of the manuscript.

FUNDING STATEMENT

This research was carried out with the support of Universitas Teknologi Akba Makassar.

COMPETING INTEREST

All authors in this study state that there are no personal interests, financial interests, and any involvement that can influence this research report.

REFERENCES

- [1] D. Song and X. Guo, "Research of the Influencing Factors of China's Foreign Exchange Reserves Based on Multiple Linear Regression Analysis Model," *Proceedings - 2020 16th Dahe Fortune China Forum and Chinese High-Educational Management Annual Academic Conference, DFHMC 2020*, pp. 180–183, 2020.
- [2] M. Liang, S. Wu, X. Wang, and Q. Chen, "A Stock Time Series Forecasting Approach Incorporating Candlestick Patterns and Sequence Similarity," *Expert Systems with Applications*, vol. 205, p. 117595, nov 2022.
- [3] M. D. Stasiak, "Candlestick The Main Mistake of Economy Research in High Frequency Markets," *International Journal of Financial Studies*, vol. 8, no. 4, pp. 1–15, 2020.
- [4] C. C. Hung and Y. J. Chen, "DPP: Deep Predictor for Price Movement from Candlestick Charts," *PLoS ONE*, vol. 16, no. 6 June 2021, pp. 1–14, 2021.
- [5] N. Nuraeni, P. Astuti, O. Irnawati, I. Darwati, and D. D. Harmoko, "High Accuracy in Forex Predictions Using the Neural Network Method Based on Particle Swarm Optimization," *Journal of Physics: Conference Series*, vol. 1641, no. 1, 2020.
- [6] S. Theodoridis, *Machine learning*, 2nd ed. Academic Press, 2020, vol. 45, no. 13.
- [7] L. Mohan, J. Pant, P. Suyal, and A. Kumar, "Support Vector Machine Accuracy Improvement with Classification," *Proceedings - 2020 12th International Conference on Computational Intelligence and Communication Networks, CICN 2020*, pp. 477–481, 2020.
- [8] A. D. Achmad, "Metode Moving Average dan Metode Support Vector Machine Untuk Prediksi Variabel Meteorologi," vol. 4, no. 1, pp. 45–50, 2017.
- [9] S. Saikin, S. Fadli, and M. Ashari, "Optimization of Support Vector Machine Method Using Feature Selection to Improve Classification Results," *JISA(Jurnal Informatika dan Sains)*, vol. 4, no. 1, pp. 22–27, 2021.
- [10] P. Aggarwal and A. K. Sahani, "Comparison of Neural Networks for Foreign Exchange Rate Prediction," *2020 IEEE 15th International Conference on Industrial and Information Systems, ICIIS 2020 - Proceedings*, no. 978, pp. 415–419, 2020.
- [11] H. Hairani, A. Anggrawan, A. I. Wathan, K. A. Latif, K. Marzuki, and M. Zulfikri, "The Abstract of Thesis Classifier by Using Naive Bayes Method," in *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*, no. August. IEEE, aug 2021, pp. 312–315. [Online]. Available: <https://ieeexplore.ieee.org/document/9537006/>
- [12] R. M. and K. L.G., "Predicting Foreign Exchange using Digital Signal Processing," *British Journal of Computer, Networking and Information Technology*, vol. 4, no. 2, pp. 1–11, 2021.
- [13] M. Islam and E. Hossain, "Foreign Exchange Currency Rate Prediction Using a GRU-LSTM Hybrid Network," *Soft Computing Letters*, vol. 3, no. August 2020, p. 100009, 2021.
- [14] A. Ramadhan, I. Palupi, and B. A. Wahyudi, "Candlestick Patterns Recognition Using CNN-LSTM Model to Predict Financial Trading Position in Stock Market," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 339–347, 2022.
- [15] A. N. Puteri, A. Arizal, and A. D. Achmad, "Feature Selection Correlation-Based pada Prediksi Nasabah Bank Telemarketing untuk Deposito," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 2, pp. 335–342, 2021.
- [16] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Computer Engineering, Science and System Journal*, vol. 4, no. 1, p. 78, 2019.

-
- [17] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends,” *Neurocomputing*, vol. 408, no. xxxx, pp. 189–215, 2020.
- [18] D. A. Pisner and D. M. Schnyer, “Support Vector Machine,” in *Machine Learning*. Elsevier, 2020, pp. 101–121.
- [19] A. Alman, A. Lawi, and Z. Tahir, “Pattern Recognition of Human Activity Based on Smartphone Data Sensors Using SVM Multiclass,” 2019.
- [20] A. J. Makrufi, W. Fawwaz, and A. Maki, “Support Vector Machine with Firefly Optimization Algorithm for Apple Fruit Disease Classification,” vol. 22, no. 1, pp. 179–190, 2022.
- [21] R. Satpathy, S. N. Mohanty, S. Satpathy, T. Choudhury, and X. Zhang, *Data Analytics in Bioinformatics: A Machine Learning Perspective*. United States of America: Wiley Interscience, 2021.
- [22] A. N. Puteri, Dewiani, and Z. Tahir, “Comparison of Potential Telemarketing Customers Predictions with a Data Mining Approach using the MLPNN and RBFNN Methods,” *2019 International Conference on Information and Communications Technology, ICOIACT 2019*, pp. 383–387, 2019.
- [23] N. Wayan, S. Saraswati, I. G. Ayu, and A. Diatri, “Recognize the Polarity of Hotel Reviews using Support Vector Machine,” vol. 22, no. 1, pp. 25–36, 2022.

[This page intentionally left blank.]