RESEARCH ARTICLE | JANUARY 18 2024

Prediction of bullish and bearish candlestick signals movement on forex using random forest and multilayer perceptron

Zebe Da; Siana Halim

AIP Conf. Proc. 2951, 030008 (2024)
https://doi.org/10.1063/5.0181370

Prediction of Bullish and Bearish Candlestick Signals Movement on Forex using Random Forest and Multilayer Perceptron

Zebe Da[a)] and Siana Halim[b)]

Industrial Engineering Department, Petra Christian University
Jl. Siwalankerto 121-131, Surabaya 60236, Indonesia

[b)] Corresponding author: halim@petra.ac.id
[a)] zebeda1032@gmail.com

Abstract. This paper discusses the bullish and bearish candlestick signals using random forest (RF) and multilayer perceptron (MLP). We have two scenarios to apply. First, we used the stochastics measurements as the features of the RF and MLP. Second, we added the candlestick features into the models. In the first scenario, the accuracy rate for the random forest is 61.68%, while the MLP gets an accuracy of 64.15%. Adding the candlestick features increases the accuracy of the prediction both for the RF and the MLP. In the second scenario, the random forest improved up to 82.08%, and MLP gained 72.88% accuracy.

INTRODUCTION

The foreign exchange (forex) market is a global market for currency trading. The value of the currency exchanged is moving up and down every second. Therefore, forecasting currency trading is a challenging problem. A professional trader will consider the fundamental and technical analysis as her/his daily basis decision tools in trading. The fundamental analysis is based on macro--economic events such as tax increases, inflations, global trade, and cash distribution. At the same time, the technical analysis provides the analysis based on the historical financial data [1,2,3].

There are many technical tools to describe the market movement, for example, the time-series, moving average chart, trend chart, and candlestick. This paper is focusing on the candlestick chart. Candlestick chart is the most used chart by traders and investors in foreign or stock exchange. The candlestick chart describes the market activity from open to closed. It has two conditions, i.e., bullish when the closing price is higher than the opening price and vice versa, bearish (see [4,5] for detail). Nowadays, huge models to predict the forex market using candlestick pattern, for example, [6,7,8,9]. Candlestick also effective for analyzing the stock and financial markets [10,11].

Orquín-Serrano [6] used the adaptive candlestick and machine learning approaches, such as random forest and AdaBoost, to predict the forex market. This paper will combine the idea of Kusuma [9] and Khaidem *et al.* [12] for predicting the bullish and bearish signals of the candlestick chart. Kusuma *et al.* [9] used a deep convolutional network (CNN) and candlestick chart to predict those signals. They [9] used CNN to analyze the candlestick charts, i.e., the image of candlestick charts. In this study, we do not use the candlestick charts but only use its feature to indicate bullish or bearish signals. We also used the Khaidem *et al.*[12]' features to model the next day signal prediction. Following [6], we used the random forest and multilayer perceptron to predict the next day's signals.

METHODS

This section describes shortly methods and features that we used for modeling the bullish and bearish candlestick signal in the forex market. The flow of the proposed solution of this problem is depicted in Fig. 1. It started from

Proceedings of the 2nd International Conference on Automotive, Manufacturing, and Mechanical Engineering (IC-AMME 2021)
AIP Conf. Proc. 2951, 030008-1–030008-8; https://doi.org/10.1063/5.0181370
Published by AIP Publishing. 978-0-7354-4202-3/$30.00

collected the data, and then we apply the exponential smoothing. Simultaneously, we then models using the technical parameters approach and candlestick approach. Those two approaches were then applied to the random forest and multi-layer perceptron model. Finally, we compared the experiments' results and conclude the best solution.

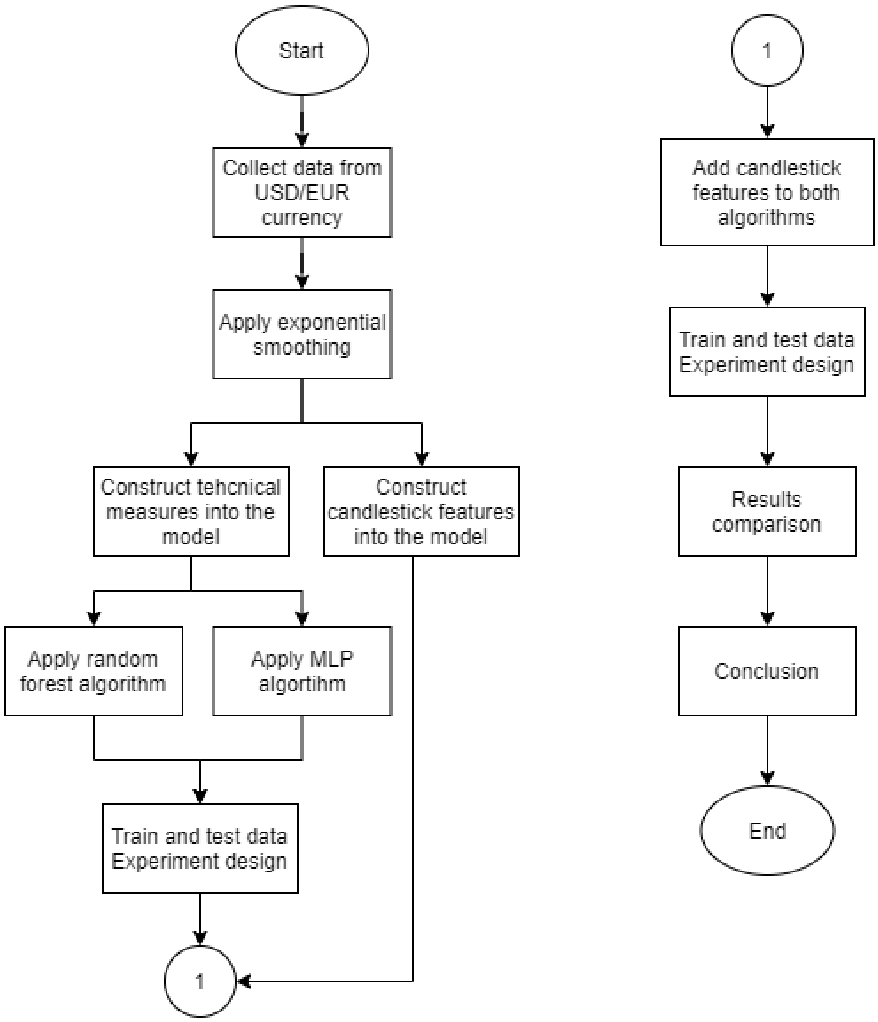

FIGURE 1. Research methodology flowchart

Candlestick Charts

A candlestick chart is like a bar chart, with a whisker at the end of the chart. Each candlestick represents four important information in the daily trading activities, i.e., the open and closed price in the thick body, high and low in the candle whisker (see [4,5] for the detail). It has two colors, which represent the price position at the closing trading date. If the closing price is higher than the opening price, then it is called a bullish candlestick; vice versa, it is called bearish (see Fig. 2).

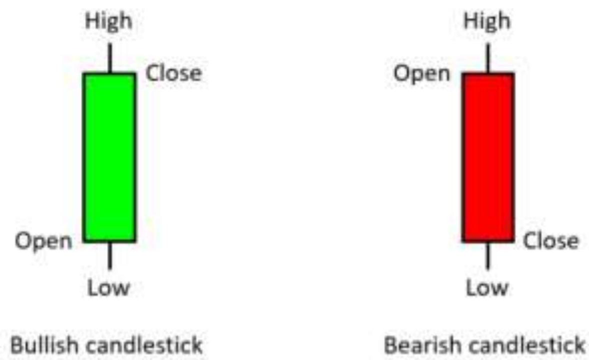

FIGURE 2. Bullish and bearish candlestick

Given four variables, that is, $open$, $high$, low and $close$, a candlestick can be represented as:

$$Body = close - open \tag{1}$$

$$BullBear = \begin{cases} 1, if\ Body > 0 \\ 0, \quad else \end{cases} \tag{2}$$

$$UpperBar = \begin{cases} high - close, if\ BullBear = 1 \\ high - open, if\ BullBear = 0 \end{cases} \tag{3}$$

$$LowerBar = \begin{cases} Low - open, if\ BullBear = 1 \\ Low - close, if\ BullBear = 0 \end{cases} \tag{4}$$

Data Preprocessing

To reduce the variance in the dataset, we then do the exponential smoothing to those four variables, as follows,

$$S_{(i)} = \alpha * Y_{(i)} + (1 - \alpha) * S_{(i-1)} \tag{5}$$

where, $Y_{(i)}\ \forall\ i = 2,3, \dots, N$ can be $open$, $high$, low, and $close$. N is the length of dataset. $S_{(0)} = Y_{(1)}$; α is the smoothing factor. In this study we used $\alpha = 0.6$.

The target to be predicted in the ith day is formulated as [6]:

$$target_i = Sign(close_{i+d} - close_i) \tag{6}$$

Technical Analysis

Following Khaidem *et al.* [12], we used the following technical analysis measurements:

Relative Strength Index (RSI)

RSI is one of the indicators to inform users of overbought and oversold conditions. Overbought happens when the price movement gets a constant rise and considered to be abnormal. Oversold occurs when the price movement gets a continuous drop and is deemed to be abnormal. It can be formulated as

$$RSI = 100 - \frac{100}{1+RS}; \tag{7}$$

$$RS = \frac{Average\ Gain\ last\ 14\ days}{Average\ Loss\ last\ 14\ days} \tag{8}$$

The value of RSI ranges from 0 to 100. The RSI indicates overbought when its' value is greater than 70 and oversold when its's value is less than 30. Average gain is the average value of 14 periods that market closed higher, and average loss is the average value of 14 periods that market closed lower.

Stochastic Oscillator

The stochastic oscillator also indicates the overbought or oversold condition. The stochastic oscillator calculation is based on 14 days of price movement. It is formulated as follows:

$$Fast\ \%K = 100 * \frac{(C - L14)}{(H14 - L14)} \tag{9}$$

$$Fast\ \%D = 3\ day\ Moving\ Average\ of\ Fast\ \%K \tag{10}$$

Where C is the current closing price; $L14$, $H14$ is the lowest and highest price on the 14 days trading windows consecutively. The stochastic oscillator has two value types, the fast and the slow. This study only used the fast-stochastic oscillator, as described in equations (7-8). The value of RSI ranges from 0 to 100. The Stochastic oscillator indicates overbought when its' value is greater than 80, and oversold when its's value is less than 20.

William %R

William %R has the same function as a stochastic oscillator. The difference only lies in the value range. William %R scales from 0 to -100. Both William %R and stochastic oscillator have the same movement in a graph. It is formulated as follows,

$$\%R = \frac{(H14 - C)}{(H14 - L14)} * -100 \tag{11}$$

The overbought signal occurs when the value is at -20 or above, while the oversold is at -80 or lower.

Moving Average Convergence Divergence (MACD)

Moving Average Convergence Divergence (MACD) is a signal indicator for a trader to buy or sell. MACD takes its value from 2 moving averages: exponential moving averages from 12 days and 26 days. Its value comes from the subtraction of 12-day EMA to 26-day EMA of the closing price.

$$MACD = EMA12(C) - EMA26(C) \tag{12}$$

$$SignalLine = EMA9(MACD) \tag{13}$$

SignalLine is an indicator for selling or buying the forex. It is the 9-day exponential moving average of MACD. When the MACD is higher than SignalLine, it is considered a buy signal and sell if the MACD value is below the SignalLine.

Price Rate of Change

The price rate of change (PROC) is the percentage of price change from the decided period. The center value of PROC is 0. If PROC is positive, it indicates the bullish signal, and vice versa, the bearish signal.

$$PROC(t) = \frac{C(t) - C(t-n)}{C(t-n)} \tag{14}$$

A review of fundamental and technical stock analysis technique can be seen e.g., in [13]

Bullish Or Bearish Signal Prediction

To predict the bullish or bearish signal (6), we used random forest and multi-layer perceptron. Random forest is an ensemble learning method for classification and regression, consisting of many decision trees [14,15,16,17] and implemented using python programming [18,19]. Multilayer Perceptron (MLP) is an artificial neural network composed of more than one perceptron. The composition of the MLP system is the same as other neural networks, which are the input layer, hidden layer, and output layer. The incoming input will pass through various functions with predetermined weights and biases [20, 21, 22, 23, 24].

RESULTS AND DISCUSSIONS

The dataset used in this study is the EUR/USD extracted from Yahoo Finance [25]. The historical data start from 1st December 2003 to 22nd May 2020. It is around 15 years daily dataset, including close, open, high, and low price. We deleted all missing values. The cleaned dataset then smoothened exponentially, using (5). We then set the features extraction (7)– (13) into a table. Since those features have different time lags, we only used the complete data set with no missing values.

In this study, we compare two scenarios using two methods, i.e., the random forest and multilayer perceptron. In the first scenario [6], we only used the technical analysis features to predict the bullish or bearish signal (6). In the second scenario, we added the candlestick charts features (1)-(4) to the first scenario.

Random Forest

First, we split the data into the training and testing set with an 80/20 proportion. The size of data training is 3388, and for the data testing, the size is 848. In the first model we want to predict the next day bullish/bearish signal, given all features in the first scenario at day $\boldsymbol{i}$th. Here, we have $\boldsymbol{d = 1}$; $\boldsymbol{target_i = Sign(close_{i+1} - close_i)}$. We let the bullish signal equal to one and the bearish signal equal to zero. Additionally, we set the trees' number in the random forest is 200 units. The model produces 61.68% accuracy. It predicts bullish 240 times, and bearish 290 times precisely (see Fig. 3).

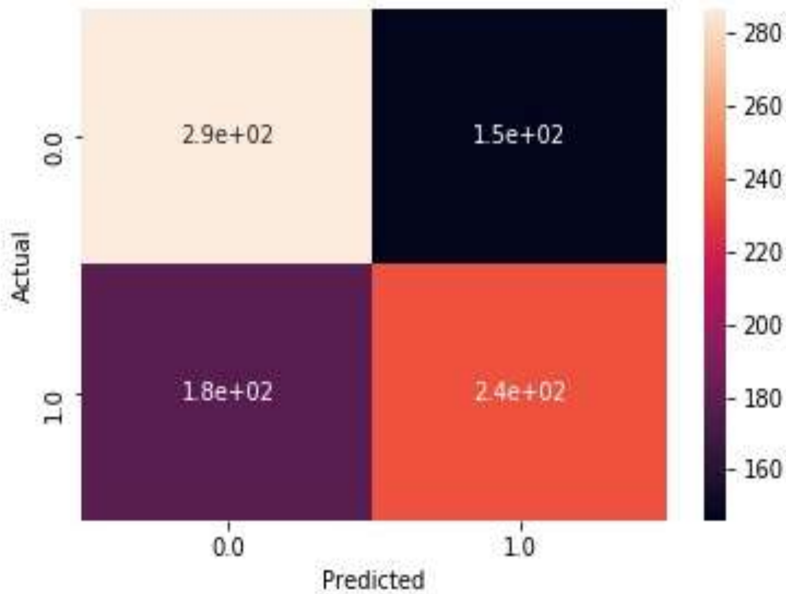

FIGURE 3. Confusion matrix of random forest next-day-prediction

TABLE 1. Accuracy, sequential accurate prediction based on the prediction lag, d for scenario 1

The prediction lag, d	Accuracy (%)	Accurate sequential prediction in days		Frequency of accurate sequential prediction
		Max.	Average	
1	61.68	12	3	115
2	61.28	13	3	106
3	60.21	11	4	100
4	56.56	17	4	87
5	58.80	17	4	85
10	51.89	19	5	66
20	56.16	32	9	47
30	59,38	50	11	44

We vary the forecast lags, i.e., the d, given all features in scenario one at the ith day. We measured how many days the prediction is accurate sequentially, as well as the accuracy percentage. We can see that the forecast at lag $d = 1$, has the highest accuracy. At $d = 1$, it can predict the signal accurately maximum 12 days sequentially, in average the signal can be predicted 3 days consecutively and it happened 115 times (see Table 1). The longer the lag, the accuracy percentage is decreasing.

Adding the candlestick features, as stated in scenario two, the accuracy is increasing from 61.68% to 79.36%. In the second scenario, we are then varying the number of candlesticks and the prediction lag d used in the prediction. The highest accuracy percentage is 82.08%. It occurred when we set $d = 1$ and the number of candlesticks used in the prediction is ten days (see Table 2).

Multilayer Perceptron

In the multilayer perceptron model, we used an experimental design approach to tune the hyperparameter of the model (see Table 3). It has the best test accuracy with two hidden layers; the number of neurons is 256 in the first layer and 64 neurons in the second layer. Moreover, we used 500 epochs. The input neurons of the multilayer perceptron are the technical analysis features (7-13), and the output is the $Target$ (6). The scaler is Min-Max scaler, activation function used in the input layers are rectified linear, while in the output layer, it used sigmoid. The optimizer used in the model is RMSprop. Figure 4 exhibits the training and testing accuracy for this model.

TABLE 2. Accuracy, sequential accurate prediction based on the prediction lag, d for scenario 2

The prediction lag d	Number of candlesticks used in the prediction	Accuracy (%)	Sequential accurate prediction in days		Frequentcy of sequential accurate prediction
			Max	Average	
1	1	79.36	29	5	112
	3	80.90	26	5	108
	5	80.90	35	5	110
	10	**82.08**	30	5	113
2	1	79.20	17	4	110
	3	74.74	26	4	117
	5	73.91	26	4	122
	10	72.61	18	4	123
5	1	60.45	15	4	103
	3	59.51	19	4	105
	5	60.45	26	4	108
	10	59.03	19	4	90
10	1	51.66	20	5	67
	3	52.60	20	5	80
	5	52.25	18	5	75
	10	52.13	17	5	67
30	1	58.67	50	12	38
	3	57.96	50	12	36
	5	57.36	50	11	36
	10	58.70	56	11	38

The second multilayer perceptron model is constructed similarly to the first multilayer perceptron model. However, in the second model, we added the candlestick features and used ten days candlestick in the input neurons. Using experiment design, we used two hidden layers, 256 and 64 neurons consecutively, with 500 epochs. The accuracy percentage is increasing from 64.15% to 72.88%.

TABLE 3. Hyperparameter tuning, experiment design

Hidden Layer	Number of neurons in each layer consecutively	Epoch	Accuracy (%)		***Run Time***
			Training	Test	
2	256, 64	500	63.87	**64.15**	3m 24s
		750	63.90	63.56	4m 58s
		1000	64.76	62.38	6m 24s
3	256, 128, 64	500	64.73	63.09	3m 43s
		750	67.27	59.42	5m 38s
		1000	67.62	59.55	7m 30s
4	512, 256, 128, 64	500	66.09	56.72	5m 42s
		750	65.82	60.61	8m 37s
		1000	66.15	58.96	11m 31s

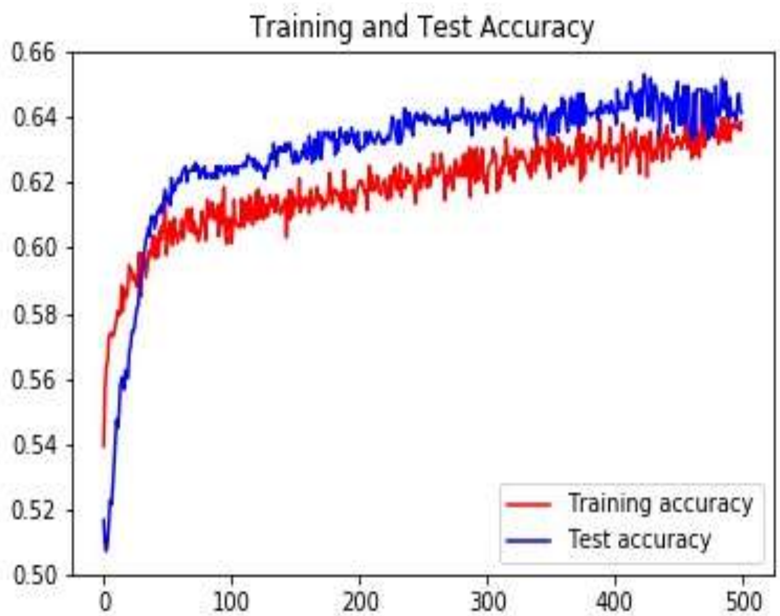

FIGURE 4. Training and testing accuracy

CONCLUSION

This paper presented the prediction of bullish and bearish signals in the foreign exchange market using random forest and multilayer perceptron model. In the first scenario, the features used in the model are only the technical features, while in the second scenario, we added the candlestick features into the model. The results show that adding candlestick features improve the accuracy of prediction. They also show that the percentage accuracy of random forest with candlestick features outperforms multilayer perceptron (MLP) model results. Table 4 exhibits the comparison of the first and second scenarios using random forest and multilayer perceptron. This model can be extended to the deep learning model, and long-short term memory to accommodate the time series nature of the foreign exchange.

TABLE 4. Results comparisons

Methods	Candlestick Features	Accuracy (%)
Random forest	Without	61.68
	With	82.08
MLP	Without	64.15
	With	72.88

REFERENCES

1. P. Chang and C. Liu, Expert Systems with Applications **34**(1), 135-144 (2008).
2. L. Menkhoff, Journal of Banking & Finance **34**(11), 2573-2586 (2010).
3. M. J. Moore and M. J. Roche, Journal of International Economics **58**(2), 387–411 (2015).
4. S. Nison, *Japanese Candlestick Charting Technique* 2nd Eds. (Prentice Hall Press. Inc., 2001)
5. G. L. Morris, *Candlestick Charting Explained: Timeless Technique for Trading Stocks and Futures* 3rd Ed (McGraw-Hill, New York, NY, 2006)
6. I. Orquín-Serrano, Mathematica **8**(5), 802-835 (2020).
7. Y. Udagawa, "Design and Implementation of Candlestick Chart Retrieval Algorithm for Predicting Stock Price Trend", in *ALLDATA 2018: In the Fourth International Conference on Big Data, Small Data, Linked Data and Open Data*, 2018, pp. 19-25.
8. T. H. Lu, Y. M, Shiu, and T. C. Liu, Review of Financial Economics **21**(2), 63-68 (2012).
9. R. M. I. Kusuma, T. T. Ho, W. C. Kao, Y.Y. Ou, and K. L. Hua, arXiv: 1903.12258v1[q-fin.GN] (2019).

10. H. A. do Prado, E. Ferneda, L. C. R. Morais, A. J. B. Luiz, and E. Matsura, Procedia Computer Science **22**, 1136-1145 (2013).
11. P. Tharavanji, V. Siraprapasiri, and K. Rajchamaha, "Profitability of candlestick charting patterns in the stock exchange of Thailand", *Sage Open*, October-December: 1-18 (2017) https://journals.sagepub.com/doi/10.1177/2158244017736799
12. L. Khaidem, S. Saha, and S.R. Dey, arXiv: 1605,00003v1[cs.LG] (2016) https://arxiv.org/abs/1605.00003
13. V. Drakopoulou, Journal of Stock & Forex Trading **5**(1), (2016)
14. M. Belgiu and L. Drăguţ, ISPRS Journal of Photogrammetry and Remote Sensing **114**, 24-31 (2016).
15. Z. Tan, Z. Yan, and G. Zhu, Heliyon **5**(8) e02310 August (2019).
16. L. Breiman, Machine Learning **45**(1), 5-32 (2001).
17. S. Mabu, M. Obayashi, and T. Kuremoto, Applied Soft Computing **36**(C) (2015).
18. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*: *Concepts, Tools and Technique to Build Intelligent Systems*, 2nd Editions, (O'Reilly, 2019).
19. W. Koehrsen, "An implementation and explanation of the random forest in python". *Medium*, 2018, retrieved from https://towardsdatascience.com/an-implementation-and-explanation-of-the-random-forest-in-python-77bf308a9b76 on 17 April 2020.
20. L. Di Persio and O. Honchar, International Journal of Circuits, Systems and Signal Processing **10**, 403-413 (2016).
21. P. Czekalski, M. Niezabitowski, and R., Styblinski, R., "ANN for FOREX forecasting and trading," *Proceedings - 2015 20th International Conference on Control Systems and Computer Science*, CSCS 2015, 322–328 (2015).
22. A. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, Journal of Economics, Finance and Administrative Science **21(41)**, 89-93 (2016).
23. A.V. Devados, T.A.A. Ligori, International Journal of Computing Algorithm **2**(2), 440-449 (2013).
24. K. S. Putri, S. Halim, International Journal of Industrial Optimization **1**(2), 71-80 (2020).
25. Yahoo Finance available at finance.yahoo.com