The experiment of text – number combination forecasting

F Rahutomo\textsuperscript{1,*}, M M Huda\textsuperscript{2}, R A Asmara\textsuperscript{1}, A Setiawan\textsuperscript{2} and A A Septarina\textsuperscript{2}

\textsuperscript{1} Information Technology Department, State Polytechnic of Malang. Jl. Soekarno-Hatta No 9, Malang 65141, Indonesia
\textsuperscript{2} Electrical Engineering Departement, State Polytechnic of Malang. Jl. Soekarno-Hatta No 9, Malang 65141, Indonesia

*faisal@polinema.ac.id

Abstract. In foreign exchange money trading, historical data are publicly available continuously. This historical data such as opening, highest, lowest, and closing rate are important variable to predict the future of rates movement. The available data is not only historical trading itself, but also from news release and expert analysis from expert trader. This kind of data contains text and number. This paper proposes in forecasting the rates by combining text and number data. The combination of text mining technique with several time series method i.e: simple moving average, weighted moving average and exponential moving average. Research period for this experiment is between 1\textsuperscript{st} December 2018 and 31\textsuperscript{st} January 2019. The currency pair are EUR – USD, USD – JPY and EUR – JPY. Forecasting results with some time series method were compared with combined time series forecasting method and naïve bayes classifier. The experiment results show that combined time series method with naïve bayes classifier delivered better accuracy level.

1. Introduction
With six trillion dollar market, forex exchange become world largest financial market which not less than 45% transaction volume comes from retail customer [1].Therefore, forecasting technique is highly sought by trader. As one of frequently used technical analysis method, time series analysis comes with many variation, but the purpose remain the same [2]. Most basic of time series analysis is simple moving average or as known as moving average, where each data consider the same value. One step above simple moving average is weighted moving average, in this method, each data is given weighted factor for every time series dataset. Another version of time series analysis is exponential moving average, which uses an exponential function as the basis in forming the weighting factor. Those three of time series method have been widely used in several researches [3-5].

In this study, we will be comparing three conventional time series methods, i.e. moving average, exponential moving average and smoothed moving average and combined text mining with three of conventional time series method in forex forecasting. EUR/USD, USD/JPY and EUR/USD are three major currency transaction that will be used, also three forecast error criteria for error measurement, i.e. mean square error (MSE), mean absolute percentage error (MAPE), and mean absolute scaled error (MASE).
2. Time series methods

2.1. Moving average
In time series method, moving average (MA) is most basic form in time series forecasting. Every data in this point is equally weighted, so there is no weighting factor applied to any of data points. As Ellis and Parbery describe in their paper [6], MA can be presented by:

\[ \bar{P}_{SM} = \frac{P_m + P_{m-1} + \cdots + P_{m-(n-1)}}{n} \]

(1)
\[ = \frac{1}{n} \sum_{i=0}^{n-1} pM - 1 \]

(2)

2.2. Weighted moving average
Weighted moving average is giving greater weight data than the older ones, the weighting factor are calculated from the sum of the day used in time series data, also known as the ‘sum of digit’ [7].

\[ WMAm = \frac{n pM + (n-1)pM - 1 + \cdots + 2p(M-n+2) + p(M-n+1)}{n + (n-1) + \cdots + 2 + 1} \]

(3)

Where \( n \) in equation above is refer to span number or period of forecasting formula and \( P_m \) refer to the value of time series data at point \( m \).

2.3. Exponential moving average
Exponential moving average gives less and less weight to data as they get older and older [8]. EMA for time series \( Y \) can be calculated recursively [9].

\[ S_t = \begin{cases} Y_1, & t = 1 \\ \alpha \cdot Y_t + (1 - \alpha) \cdot S_{t-1}, & t > 1 \end{cases} \]

(4)

Where \( Y_t \) is the value at time period \( t \), \( S_t \) is EMA value at time \( t \), and \( \alpha \) represents the degree also known as constant smoothing factor with a range value between 0 and 1. As suggested [10], we can estimate \( \alpha \) as:

\[ \alpha = \frac{2}{n+1} \]

(5)

3. Naïve bayes classifier for sentiment of text
In this study, the naïve bayes algorithm was used to get text sentiment, as one of popular machine learning algorithm. From figure 1, training data is processed with machine learning algorithm then processed in classifier along new data. Output from these processes is text sentiment. Naïve bayes (NB) known as one of most popular tool for classification, due its simplicity, high computational efficiency, and good classification accuracy. Especially for high dimensional data such as text [11]. The formula of naive bayes is expressed as:

\[ P(c|x) = \frac{P(x|c)P(c)}{P(x)} \]

(6)

Form equation above, \( x \) is word, \( c \) is category, \( P(x|c) \) is word probability in category \( c \), \( P(x) \) is probability of word, and \( P(c) \) is probability of category.

\[ P(c|X) = P(x_1|c) \times P(x_2|c) \times P(x_n|c) \times P(c) \]

(7)
4. Combining moving average methods and naïve bayes classifier

After get sentiment from naïve bayes classifier process, and the number of the sentiment from time series forecasting obtained, the next step is filtering text data and number data that have same date and same hour, text data is obtained from https://www.investing.com/analysis/forex started from 1st December 2018 until 31st January 2019, filtered by: technical, fundamental and signal. For numeric data, obtained from trading software, metatrader 4. In tab history center, we can download data trade form 1st December 2018 until 31st January 2019. After filtering both data, we use 54 data from EUR/JPY, 195 data EUR/USD, 77 data USD/JPY where each currency contain open, high, low and close data trade. Total numeric data from three currencies and four condition is 1,304 data record. After get a sentiment from text and number, then we combine those two into Table 1. For example, if predicted number have 1 or positive, and text have 1 or positive, from table above sum result is 2, it means decision is positive and degree is full. The formula to combining text and number can be expressed in Equation 8. Where \( C \) is rise or fall constant. \( N_t \) is rates or price of previous time, and \( N_{t-1} \) is rates or price in 2nd previous \( N_t \). Then, finally the combined rates prediction is calculated by equation 9.

\[
C = N_t - N_{t-1} \tag{8}
\]

\[
Prediction = N_t + (C \times Decision \times Degree) \tag{9}
\]

| Text     | Number   | Text + number | Decision | Degree |
|----------|----------|---------------|----------|--------|
| Positive | Positive Trend | 1 | 1 | 2 | Positive | Full (1) |
| Stagnant | 1 | 0 | 1 | Positive | Half (0,5) |
| Negative trend | 1 | -1 | 0 | Stagnant | 0 (0) |
| Neutral  | Positive Trend | 0 | 1 | 1 | Positive | Half (0,5) |
| Stagnant | 0 | 0 | 0 | Stagnant | 0 (0) |
| Negative trend | 0 | -1 | -1 | Negative | Half (0,5) |
| Negative | Positive Trend | -1 | 1 | 0 | Stagnant | 0 (0) |
| Stagnant | -1 | 0 | -1 | Negative | Half (0,5) |
| Negative trend | -1 | -1 | -2 | Negative | Full (1) |

Table 2. Example of formula combining text and number.

| No | \( N_t \) | \( N_{t-1} \) | \( C \) | Decision | Degree | Prediction |
|----|---------|---------|------|----------|--------|------------|
| 1  | 1.13545 | 1.13515 | 0.00031 | Positive (+) | 1 | 1.13564 |
| 2  | 1.13471 | 1.13470 | 0.00001 | Positive (+) | 0.5 | 1.13521 |
| 3  | 1.13680 | 1.13680 | 0.00000 | Stagnant (0) | 0 | 1.13680 |
| 4  | 1.13798 | 1.13745 | -0.00053 | Negative (-) | 0.5 | 1.13748 |
| 5  | 1.13820 | 1.13937 | -0.00117 | Negative (-) | 1 | 1.13920 |
From Table 2, in number 1, where data taken from EUR/USD December 5th 2019, previous rate from open is 1.13545, 2nd previous is 1.13515, C is 0.00031, decision if positive and degree is full. Than we can predict next price is 1.13645, where degree full can be assumed rate or price will be rising by 0.00100 or 100 point. In example number 2, Table 2, decision and degree is positive half, where degree half can be assumed rate or price will be rising by 0.00050 or 50 point. In number 3 decision and degree is stagnant, then prediction price won’t be changing or as same as Nt. In Table 2, example number 4, decision is negative and degree is half, rate or price is predicted down, and we can be assumed that rate or price down by 0.00050 or 50 point. And in number 5, where Nt = 1.1380, Nt - 1 = 1.13937, C = -0.00117, decision is negative and degree is full than rate or price predict decision will be down, and caused degree is full, we can have assumed rate or price will be down by 0.00100 or 100 point.

5. Forecast error measurement

In forecast error measurement, used three technique of measurement, i.e. mean absolute deviation error, mean square error, and mean absolute percentage error. Mean absolute deviation measurement the accuracy of the prediction by averaging absolute value of error. Measure absolute deviation error is use full when measuring error prediction the original series in the same unit [12]. Mean absolute percentage error is calculated using the absolute error in each period divided by the observed values that are evident for that period. Then, averaging those fixed percentage [13]. This error measurement are averages of the square of error sum between actual data and forecasted data. Tables below are comparison between time series method and time series method combined with naïve bayes classifier in three major pair i.e. EUR/JPY, EUR/USD, USD/JPY in each open, high, low and close. Every rates of price are comparison in three-time series method. Between moving average and combined moving average – naïve bayes classifier.

\[
MAD = \frac{1}{n} \sum |y_t - y_t'| 
\]

\[
MAPE = \frac{1}{n} \sum \frac{|y_t - y_t'|}{y_t} \times 100 \%
\]

\[
MSE = \frac{1}{n} \sum e_t^2
\]

Table 3. Comparison of EUR-JPY forecasting results.

| EUR JPY | Open | High | Low | Close |
|---------|------|------|-----|-------|
| MA     | 0.17919623 | 0.15608302 | 0.15922642 | 0.12687925 |
| MA + NB| 0.05566038 | 0.04245283 | 0.05188679 | 0.04433962 |
| MA     | 0.00143129 | 0.00423799 | 0.00127361 | 0.0010922 |
| MA + NB| 0.00043379 | 0.00032865 | 0.00040430 | 0.00034366 |

Table 4. Comparison of EUR-USD forecasting results.

| EUR USD | Open | High | Low | Close |
|---------|------|------|-----|-------|
| MA     | 0.00014756 | 0.000147355 | 0.000147355 | 0.000157409 |
| MA + NB| 0.00051795 | 0.00045385 | 0.00045385 | 0.00053590 |
| MA     | 0.000495283 | 0.000363208 | 0.000457547 | 0.002361781 |
| MA + NB| 0.00043393 | 0.00043393 | 0.00043393 | 0.000382076 |
| MA     | 0.00143129 | 0.00032865 | 0.00040430 | 0.00034366 |
| MA + NB| 0.00043379 | 0.00032865 | 0.00040430 | 0.00034366 |
### Table 5. Comparison of USD-JPY forecasting results.

|     | Open | High | Low   | Close  |
|-----|------|------|-------|--------|
| MADE | 0.16785455 | 0.12951429 | 0.03869104 | 0.14822857 |
| MSE  | 0.05162028  | 0.03169132  | 0.00279221  | 0.03586186  |
| MAPE | 0.00152090  | 0.00036181  | 0.00034988  | 0.00133818  |

### Table 6. Comparison of EUR-JPY forecasting results.

|     | Open | High | Low   | Close  |
|-----|------|------|-------|--------|
| MADE | 0.13715256 | 0.11751667 | 0.04723077 | 0.11791410 |
| MSE  | 0.04057526  | 0.02541692  | 0.00343241  | 0.02447997  |
| MAPE | 0.00109415  | 0.00045805  | 0.00037356  | 0.00041831  |

### Table 7. Comparison of USD-JPY forecasting results.

|     | Open | High | Low   | Close  |
|-----|------|------|-------|--------|
| MADE | 0.00108640 | 0.00051295 | 0.00010892 | 0.00011574 |
| MSE  | 0.00000217  | 0.00000037  | 0.00000024  | 0.00000022  |
| MAPE | 0.00009525  | 0.00004988  | 0.00003801  | 0.00010164  |

### Table 8. Comparison of EUR-JPY forecasting results.

|     | Open | High | Low   | Close  |
|-----|------|------|-------|--------|
| MADE | 0.13690578 | 0.09727911 | 0.05000000 | 0.11414756 |
| MSE  | 0.03641991  | 0.01653702  | 0.00250000  | 0.02202316  |
| MAPE | 0.00124062  | 0.00087534  | 0.00045019  | 0.00101648  |

### Table 9. Comparison of EUR-JPY forecasting results.

|     | Open | High | Low   | Close  |
|-----|------|------|-------|--------|
| MADE | 0.14996038 | 0.10049434 | 0.05283019 | 0.11705094 |
| MSE  | 0.03600002  | 0.01542998  | 0.00471698  | 0.02154739  |
| MAPE | 0.00119393  | 0.00079536  | 0.00041872  | 0.00103085  |

### Table 10. Comparison of EUR-JPY forecasting results.

|     | Open | High | Low   | Close  |
|-----|------|------|-------|--------|
| MADE | 0.00139692 | 0.00044872 | 0.00012789 | 0.00013305 |
| MSE  | 0.00000370  | 0.00000034  | 0.00000037  | 0.00000035  |
| MAPE | 0.00122591  | 0.00039346  | 0.00011202  | 0.00016169  |

### Table 11. Comparison of EUR-JPY forecasting results.

|     | Open | High | Low   | Close  |
|-----|------|------|-------|--------|
| MADE | 0.15508442 | 0.10295065 | 0.05324675 | 0.12401299 |
| MSE  | 0.04513844  | 0.01760311  | 0.00415864  | 0.03185608  |
| MAPE | 0.00140271  | 0.00044512  | 0.00092707  | 0.00112213  |
6. Conclusion

In this study, we know that time series method has a better error measurement when combined with naïve bayes classifier. For example, in EUR/JPY Moving average open, have mean absolute deviation error 0.001447559, mean square error 0.082453225, and mean absolute percentage error 0.001431293 compared with EUR/JPY moving average combined with mean absolute deviation error 0.055660377, have mean square error 0.004952830 and mean absolute percentage error 0.000433787. as shown in figure 2, where smallest error value is better.

Acknowledgments

The authors would like to thank State Polytechnic of Malang for supporting this research.

References

[1] Ni L, Li Y, Wang X, Zhang J, Yu J and Qi C 2019 Forecasting of Forex Time Series Data Based on Deep Learning Procedia computer science 147 647-652
[2] Zhu Y and Zhou G 2009 Technical analysis: An asset allocation perspective on the use of moving averages Journal of Financial Economics 92(3) 519-544
[3] Kapgate D 2014 Weighted moving average forecast model based prediction service broker algorithm for cloud computing International Journal of Computer Science and Mobile Computing 3(2) 71-79
[4] Raudys A, Lenčiauskas V and Malčius E 2013 Moving averages for financial data smoothing International Conference on Information and Software Technologies 34-45
[5] Hui X F and Wu Y J 2012 Research on simple moving average trading system based on SVM 2012 International Conference on Management Science & Engineering 19th Annual Conference Proceedings 1393-1397
[6] Ellis C A and Parbery S A 2005 Is smarter better? A comparison of adaptive, and simple moving average trading strategies Research in International Business and Finance 19(3) 399-411
[7] Hansun S 2013 A new approach of moving average method in time series analysis 2013 Conference on New Media Studies (ConMedia) 1-4
[8] Hunter J S 1986 The exponentially weighted moving average Journal of quality technology, 18(4) 203-210
[9] Hansun S 2016 FX forecasting using B-WEMA: Variant of Brown's Double Exponential Smoothing 2016 International Conference on Informatics and Computing (ICIC) 262-266
[10] Johnston F R, Boyland J E, Meadows M snd Shale E 1999 Some properties of a simple moving average when applied to forecasting a time series Journal of the Operational Research Society 50(12) 1267-1271
[11] Lam P, Wang L, Ngan H Y, Yung N H and Yeh A G 2017 Outlier detection in large-scale traffic
data by Naïve Bayes method and Gaussian mixture model method *Electronic Imaging* **2017**(9) 73-78

[12] Moon Y and Yao T 2011 A robust mean absolute deviation model for portfolio optimization *Comput. Oper. Res.* **38**(9) 1251–1258

[13] Khair U, Fahmi H, Al Hakim S and Rahim R 2017 Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error *J. Phys. Conf. Ser.* **930**(1)